



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 99145

TO: Sheridan Swope
Location: cm1/12D12/10D01
Art Unit: 1652
Friday, July 18, 2003

Case Serial Number: 09/966880

From: Barb O'Bryen
Location: Biotech-Chem Library
CM1-6A05
Phone: 308-4291 *1337*

barbara.obryen@uspto.gov

Search Notes

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 21:03:07 ; Search time 3550 seconds
(without alignments)
12856.023 Million cell updates/sec

Title: US-09-966-880A-7
Perfect score: 2818
Sequence: 1 agagaccatcattatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_nam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	777	27.6	953	14	BQ065440
2	760	27.0	856	9	AL559877
3	750	26.6	872	12	BG758510
4	730	25.9	743	12	BG686133
5	718	25.5	1052	14	BQ055935
6	663	23.5	664	12	BG754140

7	608	21.6	820	12	BG755526
8	606	21.5	820	12	BG757089
9	592	21.0	693	12	BF975096
10	558	19.8	670	12	BG341546
11	540	19.2	541	12	BF238155
12	523	18.6	570	10	AW504807
13	497	17.6	693	12	BG757392
14	456	16.2	889	12	BG686876
15	448	15.9	604	10	AW978582
16	448	15.9	604	10	AW978582
17	442	15.7	442	12	BF975166
18	380	13.5	511	12	BG170756
19	369	13.1	374	10	AW401901
20	353	12.5	353	9	AA831307
21	349	12.4	476	10	AW135547
22	346	12.3	464	10	AW452648
23	321	11.4	321	9	AI468242
24	312	11.1	413	10	AW768364
25	308	10.9	363	9	AI249134
26	297	10.5	956	12	BF664355
27	296	10.5	422	10	AW263139
28	296	10.5	528	9	AA954956
29	281	10.0	332	9	AA504653
30	279	9.9	928	12	BG398364
31	278	9.9	454	10	AW134750
32	207	7.3	939	12	BG341819
33	205	7.3	517	12	BG170824
34	199	7.1	215	10	AW444889
35	180	6.4	463	9	AA521498
36	153	5.4	507	9	AL581406
37	147	5.2	268	9	AA879422
38	101	3.6	1140	12	BF664352
39	85	3.0	364	10	AW182255
40	73	2.6	154	9	AA074342
41	73	2.6	878	12	BE988834
42	72	2.6	586	10	AV719004
43	70	2.5	252	10	BE008586
44	70	2.5	253	10	BE012103
45	70	2.5	318	9	AA548692
46	70	2.5	334	10	AV759459
47	70	2.5	346	9	AA211885
48	70	2.5	346	10	AV647070
49	70	2.5	437	17	AQ056722
50	70	2.5	448	14	N22395
51	70	2.5	544	17	AQ284662
52	70	2.5	642	10	AW973278
53	70	2.5	649	17	AQ741937
54	70	2.5	705	17	AG185819
55	69	2.4	571	17	BI4085
56	67	2.4	300	17	AQ276581
57	67	2.4	397	17	AZ303190
58	67	2.4	483	13	BI023361
59	67	2.4	725	17	AQ478179
60	67	2.4	841	17	AQ750497
61	66	2.3	286	10	AW882256
62	66	2.3	286	10	AW882261
63	66	2.3	286	10	AW882265
64	66	2.3	286	10	AW882267
65	66	2.3	388	10	AW518030
66	66	2.3	395	17	B56190
67	66	2.3	402	9	AA679794
68	66	2.3	414	17	B72122
69	66	2.3	430	17	AQ111323
70	66	2.3	434	17	AQ878262
71	66	2.3	451	9	AI653493
72	66	2.3	526	10	BE301068
73	65	2.3	395	10	BE156416
74	65	2.3	400	10	BE156526
75	65	2.3	443	12	BE940056
76	65	2.3	573	17	AQ318054
77	64	2.3	468	12	BG059756
78	64	2.3	470	17	AQ344499
79	64	2.3	479	14	N49425

BG755526	602716206
BF975089	602715124
BF975096	60245679
BG341546	602463652
BF238155	601811880
AW504807	UI-HF-BN0
BG757392	602711022
BG686876	602650861
AW978582	EST390691
BF975166	602444657
AI016902	ou31603.x
BG170756	602323871
AW401901	UI-HF-BK0
AA831307	oc70a06.s
AW135547	UI-H-B11-
AW452648	UI-H-B13-
AI468242	tg35905.x
AW768364	hk73d04.x
AI249134	qh79c07.x
BF664355	602146013
AW263139	xn78d03.x
AA954956	qp24a12.s
AA504653	aa63f11.s
BG398364	602439832
AW134750	UI-H-B11-
BG341819	602463552
BG170824	602323763
AW444889	UI-H-B13-
AA521498	aa73h06.s
AL581406	AL581406
AA879422	o391c11.s
BF664352	602146010
AW182255	xj72f10.x
AA074342	em15803.s
BE988834	601682124
AV719004	AV719004
BE008586	RC5-BN015
BE012103	RC5-BN105
AA548692	nk03q01.s
AV759459	AV759459
AA211885	zr90c03.s
AV647070	AV647070
AQ056722	CIT-HSP-2
N22395	yw37f08.s1
AQ284662	RPC111-78
AW973278	EST385376
AQ741937	HS-3569_B
AG185819	Pan trogl
BI4085	A-837A4.TP
AQ276581	CITBI-EI-
AZ303190	UP_436-15
BI023361	CM4-MT024
AQ478179	RPC1-11-2
AQ750497	HS-5576_B
AW882256	RC5-OT005
AW882261	RC5-OT005
AW882265	RC5-OT005
AW882267	RC5-OT005
AW518030	xx65h01.x
B56190	CIT-HSP-200
AA679794	ag92b12.s
B72122	CIT978SK-A-
AQ111323	CIT-HSP-2
AQ878262	HS-2272_B
AI653493	tg95a01.x
BE301068	ba84g06.x
BE156416	QVO-HT036
BE156526	QVO-HT036
BE940056	RC3-UT003
AQ318054	RPC111-11
BG059756	na533c05
AQ344499	RPC111-13
N49425	yv21c11.r1

226	55	2.0	708	17	AG11119 Pan trogl	299	53	1.9	583	10	BE078293	BE078293 CM0-BT061
227	55	2.0	750	17	AQ745035 HS_5506_A	300	53	1.9	617	17	AQ488260	AQ488260 RPCI-11-2
C 228	55	2.0	1069	13	BM464113 AGENCOURT	C 301	53	1.9	628	12	BG433506	BG433506 602497939
C 229	54	1.9	179	9	AA578326 nl38g04.s	302	53	1.9	632	12	BE902017	BE902017 601676824
C 230	54	1.9	219	9	AA484658 ne93g08.s	303	53	1.9	668	17	AQ505772	AQ505772 RPCI-11-2
C 231	54	1.9	251	17	AQ474929 CITBI-El-	304	53	1.9	674	17	AG132336	AG132336 Pan trogl
C 232	54	1.9	259	17	B75724 RPCI11-12C1	305	53	1.9	691	10	AV682637	AV682637 AV682637
C 233	54	1.9	273	9	A1758981 ty96b01.x	306	53	1.9	693	17	AG116744	AG116744 Pan trogl
C 234	54	1.9	300	17	AQ473028 CITBI-El-	307	53	1.9	1093	13	BM469364	BM469364 AGENCOURT
C 235	54	1.9	312	9	AA573067 nm34g08.s	C 308	52	1.8	174	9	AA369200	AA369200 EST80751
C 236	54	1.9	346	9	AA662598 nr17c11.s	309	52	1.8	200	14	H67663	H67663 yr76b07.s1
C 237	54	1.9	356	9	A1277454 qm84d07.x	310	52	1.8	221	14	H67708	H67708 yr72b07.s1
C 238	54	1.9	377	17	AQ166130 HS_3071_A	311	52	1.8	245	12	BF987027	BF987027 RC1-GN002
C 239	54	1.9	389	9	A1244127 qv90a07.x	312	52	1.8	250	9	AA318652	AA318652 EST20754
C 240	54	1.9	400	9	A1679082 tu61a02.x	313	52	1.8	262	9	AA365347	AA365347 EST76110
C 241	54	1.9	408	17	AZ518006 RPCI-11-2	314	52	1.8	305	9	AA303040	AA303040 EST114070
C 242	54	1.9	420	17	AQ036829 CIT-HSP-2	C 315	52	1.8	313	9	AA297443	AA297443 EST113044
C 243	54	1.9	446	9	AA427860 zw50a11.s	C 316	52	1.8	319	12	BF197893	BF197893 7p85g10.x
C 244	54	1.9	450	17	AQ338985 HS_5021_A	C 317	52	1.8	345	9	A1689008	A1689008 tx10e07.x
C 245	54	1.9	490	17	AQ034701 CIT-HSP-2	C 318	52	1.8	362	9	A1049634	A1049634 an31e04.x
C 246	54	1.9	548	12	BG989990 HOA4-1-F-	C 319	52	1.8	389	12	BG060038	BG060038 naf47h03.
C 247	54	1.9	565	9	AL599722 DKF2P313H	C 320	52	1.8	413	17	AQ033882	AQ033882 HS_2189_B
C 248	54	1.9	584	17	AQ309633 CITBI-El-	C 321	52	1.8	414	9	A1807372	A1807372 wf47d07.x
C 249	54	1.9	589	12	BF853273 MR3-EN009	C 322	52	1.8	423	10	AW963497	AW963497 EST375570
C 250	54	1.9	649	17	AG054765 Pan trogl	C 323	52	1.8	426	10	AW647233	AW647233 he07h01.x
C 251	54	1.9	669	17	AG157455 Pan trogl	C 324	52	1.8	445	14	R88750	R88750 yp93e01.s1
C 252	54	1.9	692	17	AG104929 Pan trogl	C 325	52	1.8	457	9	AA543015	AA543015 ni55e05.s
C 253	54	1.9	707	17	AG180372 Pan trogl	C 326	52	1.8	465	14	N95099	N95099 za28a02.r1
C 254	54	1.9	720	17	AV762203 AV762203	C 327	52	1.8	466	17	AQ412292	AQ412292 RPCI-11-1
C 255	54	1.9	801	17	AQ744761 HS_5506_A	C 328	52	1.8	497	17	AQ419955	AQ419955 RPCI-11-2
C 256	54	1.9	816	17	AQ989805 HS_3122_B	C 329	52	1.8	511	17	BH609775	BH609775 HIV12H08
C 257	54	1.9	861	14	BQ231417 AGENCOURT	C 330	52	1.8	521	17	AQ431260	AQ431260 HS_5120_A
C 258	54	1.9	966	14	BQ876394 AGENCOURT	C 331	52	1.8	525	17	AQ394682	AQ394682 CITBI-El-
C 259	54	1.9	5468	17	AQ839834 260L13-C5	C 332	52	1.8	530	17	AQ357292	AQ357292 CITBI-El-
C 260	53	1.9	183	9	AA021326 ze69c11.r	C 333	52	1.8	535	17	AQ269088	AQ269088 RPCI11-74
C 261	53	1.9	203	13	BI015299 PM0-ET025	C 334	52	1.8	551	17	AQ266700	AQ266700 RPCI11-74
C 262	53	1.9	220	13	BI058821 RC4-GN007	C 335	52	1.8	552	10	BE300871	BE300871 ba78c05.x
C 263	53	1.9	234	9	AA158288 zo76f06.s	C 336	52	1.8	578	17	AQ267476	AQ267476 RPCI11-75
C 264	53	1.9	270	9	AA323439 EST26216	C 337	52	1.8	590	9	AU127566	AU127566 AU127566
C 265	53	1.9	276	10	AW833285 RCO-TT000	C 338	52	1.8	626	12	BE90239	BE90239 601508330
C 266	53	1.9	298	14	F00214 HSB80D051.S	C 339	52	1.8	635	12	BF901984	BF901984 PM4-WT020
C 267	53	1.9	299	9	AA377102 EST89626	C 340	52	1.8	656	12	BF820327	BF820327 CM0-RT000
C 268	53	1.9	306	10	AW275788 xp55b12.x	C 341	52	1.8	690	17	AG112514	AG112514 Pan trogl
C 269	53	1.9	333	9	AA468247 nc76g04.s	C 342	52	1.8	697	9	AU127332	AU127332 Pan trogl
C 270	53	1.9	339	12	BF751955 RC4-BN053	C 343	52	1.8	710	14	BQ014503	BQ014503 UI-H-ED1-
C 271	53	1.9	352	12	BF751953 RC4-BN053	C 344	52	1.8	729	10	BE270175	BE270175 601185983
C 272	53	1.9	365	12	BF751949 RC4-BN053	C 345	52	1.8	762	13	BI910261	BI910261 603069482
C 273	53	1.9	375	9	AA828970 od81f04.S	C 346	52	1.8	821	12	BE967682	BE967682 601648988
C 274	53	1.9	375	10	AW833297 RCO-TT000	C 347	52	1.8	877	14	BQ230629	BQ230629 AGENCOURT
C 275	53	1.9	379	12	BF739035 PM4-KT004	C 348	52	1.8	889	14	BQ214673	BQ214673 AGENCOURT
C 276	53	1.9	387	10	BE157400 RCO-HT037	C 349	52	1.8	935	12	BG432843	BG432843 602496051
C 277	53	1.9	387	10	BE157414 RCO-HT037	C 350	52	1.8	1003	12	BG026261	BG026261 602291610
C 278	53	1.9	389	9	AA018517 ze50f05.r	C 351	52	1.8	1238	13	BM455282	BM455282 AGENCOURT
C 279	53	1.9	389	10	BE350475 hcl4c01.x	C 352	52	1.8	1875	13	BI009270	BI009270 Homo sap1
C 280	53	1.9	390	14	H86665 yt02d02.r1	C 353	52	1.8	175	9	BI005195	BI005195 PM3-HN007
C 281	53	1.9	395	17	AQ127179 HS_3042_B	C 354	51	1.8	177	9	AA553656	AA553656 tq46d10.x
C 282	53	1.9	404	13	BI037268 MR4-TN014	C 355	51	1.8	179	9	AA094109	AA094109 c11865.se
C 283	53	1.9	408	9	AA468318 nc77a04.s	C 356	51	1.8	186	12	BF215231	BF215231 601846224
C 284	53	1.9	413	9	AA055987 zf55d06.r	C 357	51	1.8	195	12	BG058240	BG058240 nah22e10.
C 285	53	1.9	413	9	A1571161 tn85h04.x	C 358	51	1.8	214	14	F00818	F00818 HSB66E102.S
C 286	53	1.9	416	10	AW341978 hb74d10.x	C 359	51	1.8	219	9	AA224338	AA224338 zr14e05.s
C 287	53	1.9	416	13	BI052522 PM3-GN037	C 360	51	1.8	223	12	BG231402	BG231402 na142b09.
C 288	53	1.9	465	9	A1336026 tq42h03.x	C 361	51	1.8	231	14	R92608	R92608 Yq07g05.s1
C 289	53	1.9	466	17	AQ413749 RPCI-11-2	C 362	51	1.8	235	14	BE392689	BE392689 601307548
C 290	53	1.9	473	9	AA478515 zw95c12.s	C 363	51	1.8	235	14	F13678	F13678 HSFH115.SC
C 291	53	1.9	478	17	AQ545518 CITBI-El-	C 364	51	1.8	236	10	AW854958	AW854958 PM0-CF026
C 292	53	1.9	484	10	AA0405593 UI-HF-BK0	C 365	51	1.8	238	17	AQ505893	AQ505893 RPCI-11-2
C 293	53	1.9	488	14	N52274 yv47e09.s1	C 366	51	1.8	240	14	H56912	H56912 yr07d08.s1
C 294	53	1.9	512	17	AQ586138 RPCI-11-4	C 367	51	1.8	242	9	A1307579	A1307579 tb35h05.x
C 295	53	1.9	513	17	AQ682298 HS_5500_B	C 368	51	1.8	245	9	A1682665	A1682665 wc64e10.x
C 296	53	1.9	516	13	BG995743 MR2-HT16	C 369	51	1.8	247	9	A1378730	A1378730 tb91h11.x
C 297	53	1.9	521	12	BF906164 IL3-WT026	C 370	51	1.8	247	9	AT382324	AT382324 ta71d11.x
C 298	53	1.9	528	17	AQ726110 HS_5403_B	C 371	51	1.8	247	9	AT382324	AT382324 ta71d11.x

372	51	1.8	250	17	AZ755597	ev01h05.x	C 445	51	1.8	378	9	AA525753	AA525753 ni57a04.s
373	51	1.8	253	12	BF854778	RC6-FN008	446	51	1.8	378	14	BO337978	BO337978 QV2-NN107
374	51	1.8	259	13	BI015027	PM0-RT025	C 447	51	1.8	381	17	AO374918	AO374918 RC111-15
375	51	1.8	265	12	BF811807	MR2-C1018	C 448	51	1.8	384	12	BF742112	BF742112 CMI-HB000
376	51	1.8	266	9	AA314494	EST18633	C 449	51	1.8	385	14	BO693971	BO693971 1000114.H
377	51	1.8	266	10	AW079664	xc59g12.x	C 450	51	1.8	387	9	AI061313	AI061313 an32q07.x
378	51	1.8	269	12	BF854664	MR2-EN009	C 451	51	1.8	390	9	AA229935	AA229935 nc51q10.r
379	51	1.8	269	17	BF854664	MR2-EN009	C 452	51	1.8	390	10	AW847291	AW847291 RC0-CT020
380	51	1.8	271	9	AA947269	ok20a01.s	C 453	51	1.8	391	13	BG942512	BG942512 ax26b07.x
381	51	1.8	272	17	BH141089	UP-511-21	C 454	51	1.8	393	9	AI276985	AI276985 qm59a06.x
382	51	1.8	279	17	AZ755629	ev02c07.x	C 455	51	1.8	404	9	AA553497	AA553497 nk81a05.s
383	51	1.8	280	9	AI814739	wk65b06.x	C 456	51	1.8	407	9	AI753113	AI753113 zq08c11.x
384	51	1.8	282	14	F32705	HSPD25675.H	C 457	51	1.8	416	9	AA194633	AA194633 cr06f04.s
385	51	1.8	283	12	BF680898	602156158	458	51	1.8	416	17	B79034	B79034 CIT-HSP-629
386	51	1.8	284	14	BM710652	UI-E-DX0-	459	51	1.8	419	9	AI357052	AI357052 qx62c09.x
387	51	1.8	288	9	AA569474	nm38f11.s	460	51	1.8	420	12	BE745183	BE745183 601573920
388	51	1.8	288	10	BE003190	QV4-BN009	C 461	51	1.8	420	14	BE582255	BE582255 l114q08.x
389	51	1.8	292	9	AA603530	np15b10.s	462	51	1.8	421	10	AW963470	AW963470 EST373543
390	51	1.8	294	10	BE065600	RC3-BT031	463	51	1.8	423	9	AA776665	AA776665 ac79h07.s
391	51	1.8	301	14	H69351	yu19d12.s1	C 464	51	1.8	426	17	AO118480	AO118480 HS-3011.B
392	51	1.8	303	9	AI865807	wk86c05.x	465	51	1.8	430	12	BF873332	BF873332 IL5-ET011
393	51	1.8	303	9	AA601376	no16d02.s	C 466	51	1.8	433	12	BF928492	BF928492 IL2-NT020
394	51	1.8	303	10	AW881075	QV0-OT003	C 467	51	1.8	433	17	AO046754	AO046754 RPC111-35
395	51	1.8	303	17	AA490385	RPT-11-2	C 468	51	1.8	434	14	BO710197	BO710197 AGNCOURT
396	51	1.8	307	12	BF849267	CM3-EN007	C 469	51	1.8	435	9	AI829331	AI829331 wk60f09.x
397	51	1.8	307	17	B57499	CIT-HSP-200	C 470	51	1.8	435	9	AA581471	AA581471 nd17f07.s
398	51	1.8	308	9	AA715075	nv50h02.r	471	51	1.8	435	12	BF773305	BF773305 CM3-1T004
399	51	1.8	308	12	BF976855	602146501	472	51	1.8	435	12	BG469693	BG469693 602534014
400	51	1.8	309	10	BE246822	TCRAP1E51	C 473	51	1.8	436	10	AW104748	AW104748 dx56e07.x
401	51	1.8	310	10	AW068316	cn20e11.y	474	51	1.8	437	9	AL712763	AL712763 DKF7p686p
402	51	1.8	313	12	BF676352	602084530	C 475	51	1.8	437	14	R48424	R48424 yj64f05.r1
403	51	1.8	314	17	B90405	CIT-HSP-216	C 476	51	1.8	438	14	W42997	W42997 zc24d11.s1
404	51	1.8	315	9	AA608751	ae56h07.s	C 477	51	1.8	440	17	AO089740	AO089740 HS-3006.B
405	51	1.8	319	14	F29968	HSPD20151.H	C 478	51	1.8	444	12	BF026590	BF026590 601672723
406	51	1.8	321	12	BG117013	602315864	C 479	51	1.8	445	9	AI769271	AI769271 wq36a02.x
407	51	1.8	330	9	AI285651	qu80d11.x	C 480	51	1.8	445	17	B43746	B43746 HS-1058-A2-
408	51	1.8	333	14	T98037	ye30c05.s1	C 481	51	1.8	446	10	BE302658	BE302658 ba73q03.y
409	51	1.8	334	9	AA715173	nv52g08.r	C 482	51	1.8	446	14	BM833654	BM833654 K-EST0108
410	51	1.8	335	10	AW815249	QV4-ST021	C 483	51	1.8	449	9	AA599011	AA599011 ac41e10.s
411	51	1.8	336	17	AZ755602	ev01h10.x	C 484	51	1.8	449	17	AO805344	AO805344 HS-3214.A
412	51	1.8	339	9	AA639946	np12e05.s	C 485	51	1.8	450	14	BM745900	BM745900 K-EST0019
413	51	1.8	339	17	AZ756263	ev03h02.r	C 486	51	1.8	450	17	B88194	B88194 CIT-HSP-219
414	51	1.8	339	17	AZ756706	ev15a05.r	C 487	51	1.8	455	10	BE409402	BE409402 601303664
415	51	1.8	340	10	AW855736	RC1-CT027	C 488	51	1.8	456	17	AO225744	AO225744 HS-2005.B
416	51	1.8	340	17	AZ756412	ev07d10.r	C 489	51	1.8	457	9	AI134278	AI134278 DKF7p547H
417	51	1.8	343	17	AZ756878	ew02g12.x	C 490	51	1.8	457	10	BE390587	BE390587 601284940
418	51	1.8	345	9	AA693484	ah20f10.s	C 491	51	1.8	458	9	AI288696	AI288696 qm06a08.x
419	51	1.8	346	9	AI446601	tj25a11.x	C 492	51	1.8	458	10	AW073871	AW073871 xb04c06.x
420	51	1.8	351	13	BM353189	ig76g10.x	C 493	51	1.8	458	14	R60189	R60189 yhl2h11.s1
421	51	1.8	352	9	AI611533	tt65f07.x	C 494	51	1.8	459	17	B41911	B41911 HS-1054-B1-
422	51	1.8	352	9	AI612070	tt64a06.x	C 495	51	1.8	460	9	AI361063	AI361063 qy04a03.x
423	51	1.8	352	12	BF900022	CM3-MT019	C 496	51	1.8	463	9	AI332680	AI332680 qq33q07.x
424	51	1.8	353	9	AA507837	ng86d05.s	C 497	51	1.8	468	14	N63746	N63746 za18e11.s1
425	51	1.8	355	12	BG152682	na58h12.	C 498	51	1.8	469	12	BF229305	BF229305 PM4-AN009
426	51	1.8	361	9	AA779075	zj42f01.s	499	51	1.8	469	12	BF229310	BF229310 PM4-AN009
427	51	1.8	362	9	AI266003	qm46d10.x	500	51	1.8	470	17	AO562723	AO562723 HS-2079.B
428	51	1.8	362	9	AA251477	zs08d02.r							
429	51	1.8	363	9	AI367721	qv75g11.x							
430	51	1.8	363	10	AW166611	xf64g09.x							
431	51	1.8	364	9	AI205077	an04f08.x							
432	51	1.8	364	9	AI205128	an04h08.x							
433	51	1.8	365	9	AA837817	oe39h03.s							
434	51	1.8	365	10	AW752493	RCO-CT020							
435	51	1.8	365	12	BF848354	PM4-EN006							
436	51	1.8	365	17	AO085904	HS-2164.B							
437	51	1.8	366	9	AI283938	qt71e11.x							
438	51	1.8	367	10	AV760014	APC111-66							
439	51	1.8	367	17	AQ239048	RPT111-66							
440	51	1.8	370	14	M77888	EST01472.Fe							
441	51	1.8	372	12	BF822429	MR1-RT002							
442	51	1.8	373	14	F03490	HSC12A052.n							
443	51	1.8	374	9	AI023070	ow65e09.s							
444	51	1.8	376	9	AA579179	nf28d09.s							

RESULT 1.

BQ065440

LOCUS

DEFINITION

AGENCY

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

ALIGNMENTS

BQ065440 953 bp mRNA linear EST 02-APR-2002
AGENCY: NCBI
DEFINITION: AGNCOURT_6855061 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5929977
VERSION: 5', mRNA sequence.
KEYWORDS: EST.

BQ065440.1 GI:19894486

BQ065440

BQ065440

BQ065440

BQ065440

BQ065440

BQ065440

BQ065440

BQ065440

BQ065440

BQ065440

REFERENCE 1 (bases 1 to 953)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM2108 row: p column: 10
High quality sequence stop: 634.
Location/Qualifiers
1. .953
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5929977"
/clone_lib="NIH_MGC_99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
238 a 236 c 233 g 246 t

BASE COUNT
ORIGIN

Query Match 27.6%; Score 777; DB 14; Length 953;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 777; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 TTGAAGTGAATTTTCTGGCTGAGACTTGACGAGGAGCAAGAAGACACTCTGGACACC 76
|||||
Db 1 TTGAAGTGAATTTTCTGGCTGAGACTTGACGAGGAGCAAGAAGACACTCTGGACACC 60
QY 77 ACTATGACAGCCTCTTGATGAACCGGAGAAAGTTCTTTTACCAATTCAAAAATGCCGC 136
|||||
Db 61 ACTATGACAGCCTCTTGATGAACCGGAGAAAGTTCTTTTACCAATTCAAAAATGCCGC 120
QY 137 TGGGCTAAGGTCGGCGTGAGACCTACCTGTGCTAGCTAGTGAAGAGCGGTGACAGTGT 196
|||||
Db 121 TGGGCTAAGGTCGGCGTGAGACCTACCTGTGCTAGCTAGTGAAGAGCGGTGACAGTGT 180
QY 197 ACATCCTTTTCACTGACATTTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGAATTG 256
|||||
Db 181 ACATCCTTTTCACTGACATTTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGAATTG 240
QY 257 CTCCTCCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGGTGCTACCGCGTCACC 316
Db 241 CTCCTCCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGGTGCTACCGCGTCACC 300
QY 317 TGGTTCACTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGA 376
Db 301 TGGTTCACTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGA 360
QY 377 GGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGC 436
Db 361 GGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGC 420
QY 437 AGGCTGACCGGAGGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATG 496
Db 421 AAGGCTGAGCGGAGGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATG 480
QY 497 ACCTTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATCAAGAACTTTC 556
|||||

Db 481 ACCTTCAAAGATATTTTACTGCTGGAATACCTTTTGTAGAAAACCATGAAGAAGTTTC 540
QY 557 AAAGCCTGGGAAGGGCTGCATGAAATTCAGTTTCGCTCTCCACAGACAGCTTCGCGCATC 616
Db 541 AAAGCCTGGGAAGGGCTGCATGAAATTCAGTTTCGCTCTCCACAGACAGCTTCGCGCATC 600
QY 617 CTTTCCCTCCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTTCTACTTTGGGACTTTGA 676
Db 601 CTTTCCCTCCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTTCTACTTTGGGACTTTGA 660
QY 677 TAGCAACTTCCAGGAATGCACACAGCATGAAATATCTCTGCTGGAAGACAGTGGATAAAA 736
Db 661 TAGCAACTTCCAGGAATGCACACAGCATGAAATATCTCTGCTGGAAGACAGTGGATAAAA 720
QY 737 AACAGTCCCTCAAGTCTTCTCTGTTTTTATCTTCAACTCTCACTTTCTTAGAGTTT 793
Db 721 AACAGTCCCTCAAGTCTTCTCTGTTTTTATCTTCAACTCTCACTTTCTTAGAGTTT 777

RESULT 2
AL559877
LOCUS AL559877 LTI_FL011_BCl 856 bp mRNA linear EST 16-FEB-2001
DEFINITION AL559877 Homo. sapiens cDNA clone CS0DG003YB14 5 prime
, mRNA sequence.
ACCESSION AL559877
VERSION AL559877.1 GI:12905793
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 856)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Location/Qualifiers
1. .856
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DG003YB14"
/clone_lib="LTI_FL011_BCl"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/lab_host="DH10B"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-stranded cDNA was digested with Not I and cloned
into the Not I and Eco RV sites of the pCMVSPORT 6 vector.
Library was constructed by Life Technologies. Contact :
Peng Liang Life Technologies, a division of Invitrogen
9800 Medical Center Drive Rockville, Maryland 20850, USA
Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com>"
BASE COUNT 209 a 217 c 202 g 226 t 2 others
ORIGIN

Query Match 27.0%; Score 760; DB 9; Length 856;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 810; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 62 GACACTCTGGACACCACTATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTTACCAC 121
Db 1 GACACTCTGGACACCACTATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTTACCAC 60
QY 122 TTCAAAATGTCCGTGGGCTAAGGTCGGCTGAGACCTACTGTGCTACGTAGTGAAG 181
Db 61 TTCAAAATGTCCGTGGGCTAAGGTCGGCTGAGACCTACTGTGCTACGTAGTGAAG 120
QY 182 AGGCGTGACAGTGTACATCTTTTCACTGGACTTTTGGTTATCTTCGCAATAAGAACGCG 241

121 AGGCGTGACAGTCTACATCCCTTTTCACTGGACTTTGGTTATCTTCCGAATAAGACGGC 180
242 TGCACGCTGAATGCTCTTCTCGCTACATCTCGGACTCGGACTAGACCTGACCTGCCGC 301
181 TGCACGCTGAATGCTCTTCTCGCTACATCTCGGACTCGGACTAGACCTGACCTGCCGC 240
302 TGTACCGGCTCACCTGGTTTCACTCTCGGAGCCCTGCTAGGACTGTGCCGACATGTG 361
241 TGTACCGGCTCACCTGGTTTCACTCTCGGAGCCCTGCTAGGACTGTGCCGACATGTG 300
362 GCGGACTTTCTGGAGGGAAACCCAACTCAGTCTGAGGATCTTACCGCGGCGCTCTAC 421
301 GCGGACTTTCTGGAGGGAAACCCAACTCAGTCTGAGGATCTTACCGCGGCGCTCTAC 360
422 TTTCTGTAGAGCGCGCAAGGCTGAGCCGAGGGCTGCGGCGGCTGACCGCGCGGGG 481
361 TTTCTGTAGAGCGCGCAAGGCTGAGCCGAGGGCTGCGGCGGCTGACCGCGCGGGG 420
482 CAAATAGCCATCATGACCTTCAAAGATTATTTTACTGCTGGAATCTTTTGTAGAAAAC 541
421 CAAATAGCCATCATGACCTTCAAAGATTATTTTACTGCTGGAATCTTTTGTAGAAAAC 480
542 CATGAAGAAGCTTTCAAAGCCCTGGGAAGGCTGCATGAAAAATCAGTTCTCTCCAGA 601
481 CATGAAGAAGCTTTCAAAGCCCTGGGAAGGCTGCATGAAAAATCAGTTCTCTCCAGA 540
602 CAGCTTTCGCGGATCCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGATTCGT 561
541 CAGCTTTCGCGGATCCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGATTCGT 600
662 ACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCCACACGATGAAATATCTCTGCTGA 721
601 ACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCCACACGATGAAATATCTCTGCTGA 660
722 AGACAGTGATAAAAACAGTCTTCAAGTCTCTCTGTTTATTTCTTCACTCTCACT 781
661 AGACAGTGATAAAAACAGTCTTCAAGTCTCTCTGTTTATTTCTTCACTCTCACT 720
782 TTTCTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTGA 841
721 TTTCTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTGA 780
842 AAATAGAGAAGAACACAGGCTGCGCCAGG 872
781 AAATAGAGAAGAACACAGGCTGCGCCAGG 811

RESULT 3
BG758510
LOCUS
DEFINITION
602712721F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4853069 5',
mRNA sequence.
ACCESSION
BG758510
VERSION
BG758510.1 GI:14069163
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 872)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: L1698 row: i column: 06
High quality sequence stop: 836.
FEATURES
source
1..872
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4853069"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 211 a 221 c 212 g 228 t
ORIGIN
Query Match 26.6%; Score 750; DB 12; Length 872;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 53 AGGGAAGAGACACTCTGGACACCACTATGGACAGCCTCTTTCATGAACCGGAGGAAGTTT 112
Db 39 AGGGAAGAGACACTCTGGACACCACTATGGACAGCCTCTTTCATGAACCGGAGGAAGTTT 98
QY 113 CTTTACCAAATTCAAAATATGTCGCTGGCTAAGGGTCGGCGTGAGACCTTACCTGTCTCTAC 172
Db 99 CTTTACCAAATTCAAAATATGTCGCTGGCTAAGGGTCGGCGTGAGACCTTACCTGTCTCTAC 158
QY 173 GTAGTGAAGAGCGGTGACAGTCTACATCTCTTTTCACTGGACTTTGGTTATCTTCGCAAT 232
Db 159 GTAGTGAAGAGCGGTGACAGTCTACATCTCTTTTCACTGGACTTTGGTTATCTTCGCAAT 218
QY 233 AAGAAGCGGTGCGACGCTGGAATTTGCTTCTCCGCTACATCTCGGACTGGGACCTTAGAC 292
Db 219 AAGAAGCGGTGCGACGCTGGAATTTGCTTCTCCGCTACATCTCGGACTGGGACCTTAGAC 278
QY 293 CCGCGCGCTGCTACCGGCTCACCTGGTTTACCTCTCGGAGCCCTTGCCTACGACTGTGTC 352
Db 279 CCGCGCGCTGCTACCGGCTCACCTGGTTTACCTCTCGGAGCCCTTGCCTACGACTGTGTC 338
QY 353 CGACATGTGGCGGACTTTCTGCGAGGGAAACCCCACTCAGTCTGAGGATCTTTCACGCG 412
Db 339 CGACATGTGGCGGACTTTCTGCGAGGGAAACCCCACTCAGTCTGAGGATCTTTCACGCG 398
QY 413 CGCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGGCTGCACCGC 472
Db 399 CGCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGGCTGCACCGC 458
QY 473 GCGGGGTGCAAAATAGCCATCATGACCTTCAAAGATTATTTTACTGCTGGAATACTTT 532
Db 459 GCGGGGTGCAAAATAGCCATCATGACCTTCAAAGATTATTTTACTGCTGGAATACTTT 518
QY 533 GTAGAAAACCATGAAGAAGACTTTCAAAGCTGGGAGGCTGCATGAAATTCAGTTCT 592
Db 519 GTAGAAAACCATGAAGAAGACTTTCAAAGCTGGGAGGCTGCATGAAATTCAGTTCT 578
QY 593 CTCTCCACAGACTTTCGGGCGATCCTTTTGGCCCTGTATGAGGTTGATGACTTACGAGAC 652
Db 579 CTCTCCACAGACTTTCGGGCGATCCTTTTGGCCCTGTATGAGGTTGATGACTTACGAGAC 638
QY 653 GCATTTCTGACTTTTGGGACTTTTGTAGCAACTTCCAGGAATGTCCACACGATGAATAT 712
Db 639 GCATTTCTGACTTTTGGGACTTTTGTAGCAACTTCCAGGAATGTCCACACGATGAATAT 698
QY 713 CTCTCTGAAGACAGTGGATAAAAACAGTCTTCAAGTCTTCTCTGTTTATTTCTTCA 772
Db 699 CTCTCTGAAGACAGTGGATAAAAACAGTCTTCAAGTCTTCTCTGTTTATTTCTTCA 758

```

QY 773 ACTCTCACTTTCTTAGAGTTTACAGAAAAATATTATATACGACTCTTTAAAAAGATCT 832
    |||||||
Db 759 ACTCTCACTTTCTTAGAGTTTACAGCAAAATATTATATACGACTCTTTAAAAAGATCT 818
    |||||||
QY 833 ATGCTTGAAATAGAGAAGG 853
    |||||||
Db 819 ATGCTTGAAATAGAGAAGG 839

RESULT 4
BG686133 743 bp mRNA linear EST 01-MAY-2001
LOCUS 602638412F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4766234 5',
DEFINITION mRNA sequence.
ACCESSION BG686133
VERSION BG686133.1 GI:13917530
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 743)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1626 row: g column: 03
High quality sequence stop: 740.
Location/Qualifiers
1..743
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4766234"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 176 a 197 c 188 g 182 t
ORIGIN

Query Match 25.9%; Score 730; DB 12; Length 743;
Best Local Similarity 100.08; Pred. No. 0;
Matches 730; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GAACCATCATTAATTGAAGTGAGATTTTCTGGCCTGAGACTTGCGAGGAGCGCAAGAGA 63
    |||||||
Db 2 GAACCATCATTAATTGAAGTGAGATTTTCTGGCCTGAGACTTGCGAGGAGCGCAAGAGA 61
    |||||||
QY 64 CACTCTGCAGACCACCATATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTTACCAATT 123
    |||||||
Db 62 CACTCTGCAGACCACCATATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTTACCAATT 121
    |||||||
QY 124 CAAAATGTCCGCTGGGCTAAGGGTCGGCGTGAGACCTACCTGTCTAGTAGTGAAGAG 183
    |||||||
Db 122 CAAAATGTCCGCTGGGCTAAGGGTCGGCGTGAGACCTACCTGTCTAGTAGTGAAGAG 181
    |||||||

```

```

QY 184 GCGTGACAGTGCTACATCTCTTTTTCACGTGGACTTTGGTTATCTTCCGAATAAGAAGCGGTG 243
    |||||||
Db 182 GCGTGACAGTGCTACATCTCTTTTTCACGTGGACTTTGGTTATCTTCCGAATAAGAAGCGGTG 241
    |||||||
QY 244 CCAGCTGGAATTGCTCTTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTG 303
    |||||||
Db 242 CCAGCTGGAATTGCTCTTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTG 301
    |||||||
QY 304 CTACCGGCTCACCTGGTTTACCTCTCTGGAGCCCTGCTACGACTCTGCCCCGACATGTGGC 363
    |||||||
Db 302 CTACCGGCTCACCTGGTTTACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGC 361
    |||||||
QY 364 CGACTTTCTCGGAGGGAAACCCCAACCTCAGTCTGAGGATCTTCAACCGCGCCCTCTACTT 423
    |||||||
Db 362 CGACTTTCTCGGAGGGAAACCCCAACCTCAGTCTGAGGATCTTCAACCGCGCCCTCTACTT 421
    |||||||
QY 424 CTGTGAGACCGCAAGGCTGAGCCCGGAGGGGCTGGGCGGCTGCACCGCGCGGGGTGCA 483
    |||||||
Db 422 CTGTGAGACCGCAAGGCTGAGCCCGGAGGGGCTGGGCGGCTGCACCGCGCGGGGTGCA 481
    |||||||
QY 484 AATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAAACCA 543
    |||||||
Db 482 AATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAAACCA 541
    |||||||
QY 544 TGAAGAAGAACTTTCAAAGCCTGGGAAGGCTGCATGAAATTCAGTTCCTCTCCAGACA 603
    |||||||
Db 542 TGAAGAAGAACTTTCAAAGCCTGGGAAGGCTGCATGAAATTCAGTTCCTCTCCAGACA 601
    |||||||
QY 604 GCTTCGGCGCATCTTTTGGCCCTGTATGAGGTGTAGTACTTACGAGACGATTTCTGTAC 663
    |||||||
Db 602 GCTTCGGCGCATCTTTTGGCCCTGTATGAGGTGTAGTACTTACGAGACGATTTCTGTAC 661
    |||||||
QY 664 TTTGGGACTTTGTATAGCAACTTCCAGGAATGTCCACACACGATGAAATATCTTCTGCTGAAG 723
    |||||||
Db 662 TTTGGGACTTTGTATAGCAACTTCCAGGAATGTCCACACACGATGAAATATCTTCTGCTGAAG 721
    |||||||
QY 724 ACAGTGGATA 733
    |||||||
Db 722 ACAGTGGATA 731

RESULT 5
BG055935 1052 bp mRNA linear EST 29-MAR-2002
LOCUS AGENCOURT_6796291 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5808181
DEFINITION 5', mRNA sequence.
ACCESSION BG055935
VERSION BG055935.1 GI:19815262
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1052)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2051 row: m column: 14
High quality sequence stop: 665.
Location/Qualifiers
1..1052
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5808181"
FEATURES
source

```

```

/clone_lib="NIH_MGC_99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dr priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT      276 a      250 c      258 g      266 t      2 others
ORIGIN

Query Match      25.5%; Score 718; DB 14; Length 1052;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 718; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 TTGAAGTGAGATTTTCTGGCTGAGACTTGCAGGAGGCAAGAGACACTCTGGACACC 76
Db 1 TTGAAGTGAGATTTTCTGGCTGAGACTTGCAGGAGGCAAGAGACACTCTGGACACC 60

Qy 77 ACTATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTACCAATTCAAAATTCGCC 136
Db 61 ACTATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTACCAATTCAAAATTCGCC 120

Qy 137 TGGGTAAAGGTCGCGGTGAGACCTACCTGTGCTAGTAGTGAAGCGGTGACAGTGT 196
Db 121 TGGGTAAAGGTCGCGGTGAGACCTACCTGTGCTAGTAGTGAAGCGGTGACAGTGT 180

Qy 197 ACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATTG 256
Db 181 ACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATTG 240

Qy 257 CTCTTCTCCGCTACATCTCGACCTGGGACCTAGACCTTGGCGCTGTACCGCGTCAACC 316
Db 241 CTCTTCTCCGCTACATCTCGACCTGGGACCTAGACCTTGGCGCTGTACCGCGTCAACC 300

Qy 317 TGGTTACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGCGACTTCTCGGA 376
Db 301 TGGTTACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGCGACTTCTCGGA 360

Qy 377 GGGAAACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGC 436
Db 361 GGGAAACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGC 420

Qy 437 AAGGCTGAGCCGAGGGCTGCGGGCTGCACCGCGCGGGGTGCAATACCCATCATG 496
Db 421 AAGGCTGAGCCGAGGGCTGCGGGCTGCACCGCGCGGGGTGCAATACCCATCATG 480

Qy 497 ACCTTCAAGATTTATTTTACTGCTGGAAATCTTTGTAGAAAACCATGAAAGAACTTTC 556
Db 481 ACCTTCAAGATTTATTTTACTGCTGGAAATCTTTGTAGAAAACCATGAAAGAACTTTC 540

Qy 557 AAAGCCTGGGAAGGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGGGCGATC 616
Db 541 AAAGCCTGGGAAGGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGGGCGATC 600

Qy 617 CTTTTCGCCCTGTATGAGTTGATGACATTACGAGGCAATTCGATCTTTGGGACTTTGA 676
Db 601 CTTTTCGCCCTGTATGAGTTGATGACATTACGAGGCAATTCGATCTTTGGGACTTTGA 660

Qy 677 TAGCAACTTCCAGGAATGTCACACAGGATGAATATCTCTGCTGAAGACAGTGGATAA 734
Db 661 TAGCAACTTCCAGGAATGTCACACAGGATGAATATCTCTGCTGAAGACAGTGGATAA 718

```

```

RESULT 6
BG754140
LOCUS      602709681F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4846005 5',
DEFINITION mRNA sequence.

```

```

ACCESSION      BG754140
VERSION        BG754140.1
KEYWORDS       EST
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS        1 (bases 1 to 664)
              NIH-MGC http://mgc.nci.nih.gov/.
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-re@mail.nih.gov
              Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone Distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LCM1686 row: b column: 22
              High quality sequence stop: 662.
              Location/Qualifiers
                1..664
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4846005"
                /clone_lib="NIH_MGC_48"
                /tissue_type="Primary B-cells from tonsils (cell line)"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
                Site_2: EcoRI; cDNA made by oligo-dr priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCAGGAG(G). Size-selected >500bp
                for average insert size 1.8kb. Library constructed by Ling
                Hong in the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC Library."
BASE COUNT      199 a      121 c      130 g      214 t
ORIGIN

Query Match      23.5%; Score 663; DB 12; Length 664;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 663; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 634 GGTTCATGACTTACCAGAGCGCATTTTCGTACTTTTGGGACTTTTGATAGCAATTTCCAGGAAT 693
Db 2 GGTTCATGACTTACCAGAGCGCATTTTCGTACTTTTGGGACTTTTGATAGCAATTTCCAGGAAT 61

Qy 694 GTCACACAGCATGAATATCTCTGCTGAAGACAGTGGATATAAACAAGTCTTCAAGTCT 753
Db 62 GTCACACAGCATGAATATCTCTGCTGAAGACAGTGGATATAAACAAGTCTTCAAGTCT 121

Qy 754 TCTCTGTTTTTATCTTCAACTCTCCTTCTTAGAGTTTACAGAAAAAATATTTATATA 813
Db 122 TCTCTGTTTTTATCTTCAACTCTCCTTCTTAGAGTTTACAGAAAAAATATTTATATA 181

Qy 814 CGACTCTTTAAAGACATCTATCTCTTGAANAATAGAGGAACACAGGTCGCGCAGGA 873
Db 182 CGACTCTTTAAAGACATCTATCTCTTGAANAATAGAGGAACACAGGTCGCGCAGGA 241

Qy 874 CGTGTGCAATTTGGTGCAGTTTGTGAATGCAACATTTGCCCTTACTGGGAATTAACAACT 933
Db 242 CGTGTGCAATTTGGTGCAGTTTGTGAATGCAACATTTGCCCTTACTGGGAATTAACAACT 301

Qy 934 CGAGGACCTGGGAGGATCCTAAAGTGTCAACCGTTTTTCTATGACCTTTTATAGTAGATGAG 993
Db 302 CGAGGACCTGGGAGGATCCTAAAGTGTCAACCGTTTTTCTATGACCTTTTATAGTAGATGAG 361

Qy 994 AGCAAGAGGTAGATCTCTTAAAGAGCATGGTGCAGGATCAAAATGTTTTTATATCAACATCC 1053
Db 362 AGCAAGAGGTAGATCTCTTAAAGAGCATGGTGCAGGATCAAAATGTTTTTATATCAACATCC 421

```

QY 1054 TTTATTATTGATTCATTGAGTTAAACAGTGGTGTAGTGATAGATTTTCTATTCTTTT 1113
|||||
Db 422 TTTATTATTGATTCATTGAGTTAAACAGTGGTGTAGTGATAGATTTTCTATTCTTTT 481
QY 1114 CCTTCAGCTTTACTTTCAAGTAACACAACACTCTCCATCAGGCCATGATCATAGGACC 1173
|||||
Db 482 CCTTCAGCTTTACTTTCAAGTAACACAACACTCTCCATCAGGCCATGATCATAGGACC 541
QY 1174 TCCTAATGAGAGTATCTGGGTGATTGTGACCCCAACCACTCTCTCCAAAGCACTTAATATC 1233
|||||
Db 542 TCCTAATGAGAGTATCTGGGTGATTGTGACCCCAACCACTCTCTCCAAAGCACTTAATATC 601
QY 1234 CAATCATGCGCTGTATGTTTTTAATCAGCAGAAGCATGTTTTTATGTTTGTACAAAAGAAG 1293
|||||
Db 602 CAATCATGCGCTGTATGTTTTTAATCAGCAGAAGCATGTTTTTATGTTTGTACAAAAGAAG 661
QY 1294 ATT 1296
Db 662 ATT 664
RESULT 7
BG755526 820 bp mRNA linear EST 15-MAY-2001
LOCUS 602716206F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4856568 5',
DEFINITION mRNA sequence.
ACCESSION BG755526
VERSION BG755526.1 GI:14066179
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM1707 row: k column: 01
High quality sequence stop: 638.
Location/Qualifiers

FEATURES
source
1..820
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4856568"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT 245 a 161 c 168 g 246 t
ORIGIN
Query Match 21.6%; Score 608; DB 12; Length 820;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 TTGGGACTTTGTATAGCAAACTTCCAGGAATGTCCACACAGATGAAATATATCTCTGCTGAAGA 724
|||||
Db 32 TTGGGACTTTGTATAGCAAACTTCCAGGAATGTCCACACAGATGAAATATATCTCTGCTGAAGA 91
QY 725 CAGTGGATAAAAAACAGTCCCTTCAAGTCTCTCTGTTTTTATTCTTCAACTCTCACTTTTC 784
|||||
Db 92 CAGTGGATAAAAAACAGTCCCTTCAAGTCTCTCTGTTTTTATTCTTCAACTCTCACTTTTC 151
QY 785 TTAGAGTTTTACAGAAAAAATATTTATATACGACTCTTTTAAAAAGATCTATGTCTTCAAAA 844
|||||
Db 152 TTAGAGTTTTACAGAAAAAATATTTATATACGACTCTTTAAAAAGATCTATGTCTTGAANA 211
QY 845 TAGAGAAGAACACAGGCTCGCCAGGGACGTGTGCAATTTGGTGAGTGTGAGTGTGAGTGTG 904
|||||
Db 212 TAGAGAAGAACACAGGCTCGCCAGGGACGTGTGCAATTTGGTGAGTGTGAGTGTGAGTGTG 271
QY 905 CATTGTCCCTACTCGGGAATAACAGAACTGCAGGACCTGGGAGCATCCCTAAAAGTGTCAAC 964
|||||
Db 272 CATTGTCCCTACTCGGGAATAACAGAACTGCAGGACCTGGGAGCATCCCTAAAAGTGTCAAC 331
QY 965 GTTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAGGTAGATCTCTAAAAAGCATGGTGA 1024
|||||
Db 332 GTTTTTCTATGACTTTTAGGTAGGATGAGAGCAGAGGTAGATCTCTAAAAAGCATGGTGA 391
QY 1025 GAGGATCAAAATGTTTTTATATCAACATCCTTTTATTTGATTCATTTGAGTTAAAGTGTG 1084
|||||
Db 392 GAGGATCAAAATGTTTTTATATCAACATCCTTTTATTTGATTCATTTGAGTTAAAGTGTG 451
QY 1085 GTGTTAGTGATAGATTTTCTATTCTTTCCCTTGACGTTTACTTTTCAAGTAAACAACAA 1144
|||||
Db 452 GTGTTAGTGATAGATTTTCTATTCTTTCCCTTGACGTTTACTTTTCAAGTAAACAACAA 511
QY 1145 TCTTCCATCAGGCCATGATCTATAGGACCTCTCTAATGAGAGATATCTGGGTGATTTGAGCC 1204
|||||
Db 512 TCTTCCATCAGGCCATGATCTATAGGACCTCTCTAATGAGAGATATCTGGGTGATTTGAGCC 571
QY 1205 CCAAAACCATCTCTCCAAAGCATTAATATCAATCATGCGCTGTATGTTTTTAATCATCAGA 1264
|||||
Db 572 CCAAAACCATCTCTCCAAAGCATTAATATCAATCATGCGCTGTATGTTTTTAATCATCAGA 631
QY 1265 AGCATGTT 1272
Db 632 AGCATGTT 639
RESULT 8
BG757089 820 bp mRNA linear EST 15-MAY-2001
LOCUS 602715124F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4855517 5',
DEFINITION mRNA sequence.
ACCESSION BG757089
VERSION BG757089.1 GI:14067742
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM1704 row: o column: 06
High quality sequence stop: 675.
Location/Qualifiers

FEATURES


```
source
1. .820
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:485517"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      210 a   230 c   200 g   180 t
ORIGIN
Query Match      21.5%; Score 606; DB 12; Length 820;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 656; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AGAGAACCATCAATTAATGAAGTGAGATTTTCTGGCCTGAGACTTTCGAGGAGGCAAGA 60
Db 6 AGAGAACCATCAATTAATGAAGTGAGATTTTCTGGCCTGAGACTTTCGAGGAGGCAAGA 65
QY 61 AGACACTCTGGACACCACTATGGACAGCCTCTTTGATGAACCGGAGGAAGTTCTTTACCA 120
Db 66 AGACACTCTGGACACCACTATGGACAGCCTCTTTGATGAACCGGAGGAAGTTCTTTACCA 125
QY 121 ATTCAAAATGTCGCTGGCTAAGGTCGGCTGAGACTACTCTGTCTACGTAGTAA 180
Db 126 ATTCAAAATGTCGCTGGCTAAGGTCGGCTGAGACTACTCTGTCTACGTAGTAA 185
QY 181 GAGCGGTGACAGTGCTACATCTTTTCACTGGACTTTTGGTTATCTTCGCAATAAGAACGG 240
Db 186 GAGCGGTGACAGTGCTACATCTTTTCACTGGACTTTTGGTTATCTTCGCAATAAGAACGG 245
QY 241 CTCGCCAGTGGAAATGCTTCTCCGCTACATCTCGGACTGGACCTAGACCTGGCCG 300
Db 246 CTCGCCAGTGGAAATGCTTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCG 305
QY 301 CTGCTACCGGCTCACCTGGTTACCTCCTGGAGCCCTGTACGACTGTGCCCGACATGT 360
Db 306 CTGCTACCGGCTCACCTGGTTACCTCCTGGAGCCCTGTACGACTGTGCCCGACATGT 365
QY 361 GGCCGACTTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTTA 420
Db 366 GGCCGACTTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTTA 425
QY 421 CTTCTGTGAGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGCTGCACCGCGCGGGT 480
Db 426 CTTCTGTGAGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGCTGCACCGCGCGGGT 485
QY 481 GCAAAATAGCCATCATGACCTTCAAAGATTATTTTACTGTGGAATACTTTTGTAGAAA 540
Db 486 GCAAAATAGCCATCATGACCTTCAAAGATTATTTTACTGTGGAATACTTTTGTAGAAA 545
QY 541 CCATGAAGAACTTTTCAAAGCCTGGGAAGGCTGCGATGAAATTCAGTTCTGCTCTCCAG 600
Db 546 CCATGAAGAACTTTTCAAAGCCTGGGAAGGCTGCGATGAAATTCAGTTCTGCTCTCCAG 605
QY 601 ACAGCTTCGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCATT 657
Db 606 ACAGCTTCGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCATT 662
RESULT 9
LOCUS BF975096 60245679F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4336722 5',
DEFINITION mRNA sequence.
```

```
ACCESSION BF975096 GI:12342311
VERSION BF975096.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 693)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1209 row: n column: 19
High quality sequence stop: 692.
Location/Qualifiers
1. .693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4336722"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      179 a   167 c   162 g   185 t
ORIGIN
Query Match      21.0%; Score 592; DB 12; Length 693;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 592; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 304 CTACCGCGTCACCTGGTTACCTTCCTGGAGCCCTGCTAGGACTGTGCCGACATGTGGC 363
Db 1 CTACCGCGTCACCTGGTTACCTTCCTGGAGCCCTGCTAGGACTGTGCCGACATGTGGC 60
QY 364 CGACTTTTCTCGGAGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTT 423
Db 61 CGACTTTTCTCGGAGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTT 120
QY 424 CTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGCTGCACCGCGCGGGTGCA 483
Db 121 CTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGCTGCACCGCGCGGGTGCA 180
QY 484 AATAGGCATCATGACCTTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCA 543
Db 181 AATAGGCATCATGACCTTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCA 240
QY 544 TGAAGAACTTTTCAAAGCCTGGGAAGGCTGCATGAAATTCAGTTCTGCTCTCCAGACA 603
Db 241 TGAAGAACTTTTCAAAGCCTGGGAAGGCTGCATGAAATTCAGTTCTGCTCTCCAGACA 300
QY 604 GCTTTCGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCATTTCGTAC 663
Db 301 GCTTTCGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCATTTCGTAC 360
QY 664 TTTGGGACTTTTGATAGCAACTTCCAGGAATGTCACACACGATGAAATATCTCTGCTGAAG 723
Db 361 TTTGGGACTTTTGATAGCAACTTCCAGGAATGTCACACACGATGAAATATCTCTGCTGAAG 420
```



```

QY 724 ACAGTGGATAAAAAACAGCTCTCAAGTCTTCTGTTTTTATCTTCAACTCTCACITTT 783
    |||||||
Db 421 ACAGTGGATAAAAAACAGCTCTCAAGTCTTCTGTTTTTATCTTCAACTCTCACITTT 480
QY 784 CTTAGAGTTTACAGAAAAAATATTTATATACAGACTCTTTAAAAAGATCTATGCTTGAAA 843
    |||||||
Db 481 CTTAGAGTTTACAGAAAAAATATTTATATACAGACTCTTTAAAAAGATCTATGCTTGAAA 540
QY 844 ATAGAGAAGGAACACAGGCTCGGCCAGGAGCTGCTGCAATTGGTGCACTTT 895
    |||||||
Db 541 ATAGAGAAGGAACACAGGCTCGGCCAGGAGCTGCTGCAATTGGTGCACTTT 592

RESULT 10
LOCUS BG341546 670 bp mRNA linear EST 27-FEB-2001
DEFINITION 602463652F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4576548 5',
    mRNA sequence.
ACCESSION BG341546
VERSION BG341546.1 GI:13147984
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1289 row: g column: 13
High quality sequence stop: 650.
Location/Qualifiers
1..670
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4576548"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 200 a 122 c 134 g 214 t
ORIGIN
Query Match 19.8%; Score 558; DB 12; Length 670;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 558; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 TTGGGACTTTGATAGCAACTCCAGGAATGTCACACACATGAATATCTCTGCTGAAGA 724
    |||||||
Db 32 TTGGGACTTTGATAGCAACTCCAGGAATGTCACACACATGAATATCTCTGCTGAAGA 91
QY 725 CAGTGGATAAAAAACAGTCCCTTCAAGCTCTCTGTTTTTATCTTCAACTCTCACTTTC 784
    |||||||
Db 92 CAGTGGATAAAAAACAGTCCCTTCAAGCTCTCTGTTTTTATCTTCAACTCTCACTTTC 151

```

```

QY 785 TTAGAGTTTACAGAAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTTGAAA 844
    |||||||
Db 152 TTAGAGTTTACAGAAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTTGAAA 211
QY 845 TAGAGAAAGGAACACAGGCTCGCCAGGAGACGTGCTGCAATTGGTGCAGTTTTGAATGCAA 904
    |||||||
Db 212 TAGAGAAAGGAACACAGGCTCGCCAGGAGACGTGCTGCAATTGGTGCAGTTTTGAATGCAA 271
QY 905 CATTCTCCCTACTCGGGAATAACAGAACTCGAGGACCTGGGAGCATCCTAAAGTGTCAAC 964
    |||||||
Db 272 CATTCTCCCTACTCGGGAATAACAGAACTCGAGGACCTGGGAGCATCCTAAAGTGTCAAC 331
QY 965 GTTTTCTATGACTTTTAGGTAGGATGAGACAGAGGTAGATCATCTAAAAAGCATGGTGA 1024
    |||||||
Db 332 GTTTTCTATGACTTTTAGGTAGGATGAGACAGAGGTAGATCATCTAAAAAGCATGGTGA 391
QY 1025 GAGGATCAAAATGTTTTTATATCAACATCCTTTATTTATTTGATTCATTTGAGTTAACAGTG 1084
    |||||||
Db 392 GAGGATCAAAATGTTTTTATATCAACATCCTTTATTTATTTGATTCATTTGAGTTAACAGTG 451
QY 1085 GTGTTAGTGATAGATTTTCTATTCTTTTCCCTTGACGTTTACTTTCAAGTAACACAAC 1144
    |||||||
Db 452 GTGTTAGTGATAGATTTTCTATTCTTTCCCTTGACGTTTACTTTCAAGTAACACAAC 511
QY 1145 TCTTCCATCAGCCCATGATCTATAGGACCTCCTTAATGAGAGTATCTGGGTGATTTGACG 1204
    |||||||
Db 512 TCTTCCATCAGCCCATGATCTATAGGACCTCCTTAATGAGAGTATCTGGGTGATTTGACG 571
QY 1205 CCAACCATTCTCCAAA 1222
    |||||||
Db 572 CCAACCATTCTCCAAA 589

RESULT 11
LOCUS BF238155
DEFINITION 601811880F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4054915 5',
    mRNA sequence.
ACCESSION BF238155
VERSION BF238155.1 GI:11152074
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM895 row: p column: 20
High quality sequence stop: 541.
Location/Qualifiers
1..541
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4054915"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 200 a 122 c 134 g 214 t
ORIGIN
Query Match 19.8%; Score 558; DB 12; Length 670;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 558; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 665 TTGGGACTTTGATAGCAACTCCAGGAATGTCACACACATGAATATCTCTGCTGAAGA 724
    |||||||
Db 32 TTGGGACTTTGATAGCAACTCCAGGAATGTCACACACATGAATATCTCTGCTGAAGA 91
QY 725 CAGTGGATAAAAAACAGTCCCTTCAAGCTCTCTGTTTTTATCTTCAACTCTCACTTTC 784
    |||||||
Db 92 CAGTGGATAAAAAACAGTCCCTTCAAGCTCTCTGTTTTTATCTTCAACTCTCACTTTC 151

```

Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

```
BASE COUNT      120 a  150 c  142 g  129 t
ORIGIN
    Query Match      19.2%; Score 540; DB 12; Length 541;
    Best Local Similarity 100.0%; Pred. No. 0;
    Matches 540; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy  4 GAACCATCAATTAATGAAGTGAGATTTTCTGGCTGAGACTTGCGAGGAGGCAAGAAGA 63
Db  2 GAACCATCAATTAATGAAGTGAGATTTTCTGGCTGAGACTTGCGAGGAGGCAAGAAGA 61
Qy  64 CACTCTGGACACCATATGGACAGCCTCTGTATGAACCGGAGGAAGTTCTTTACCAATT 123
Db  62 CACTCTGGACACCATATGGACAGCCTCTGTATGAACCGGAGGAAGTTCTTTACCAATT 121
Qy  124 CAAAATGTCCGCTGGGCTAAGGTCGGCTGAGACCTACCTGTGCTACGTAGTGAAGAG 183
Db  122 CAAAATGTCCGCTGGGCTAAGGTCGGCTGAGACCTACCTGTGCTACGTAGTGAAGAG 181
Qy  184 GCGTGACAGTGCTACATCCTTTTCACTGGACTTTTGTATCTTCGCAATAAGAAGCGGTG 243
Db  182 GCGTGACAGTGCTACATCCTTTTCACTGGACTTTTGTATCTTCGCAATAAGAAGCGGTG 241
Qy  244 CCACCTGGAAATGCTTCCTCCGCTACATCTCGAGCTGGGACCTAGACCTGGCGGCTG 303
Db  242 CCACCTGGAAATGCTTCCTCCGCTACATCTCGAGCTGGGACCTAGACCTGGCGGCTG 301
Qy  304 CTACCGCGTCACCTGTTTCACTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCG 363
Db  302 CTACCGCGTCACCTGTTTCACTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCG 361
Qy  364 CGACTTTCTGCGAGGAAACCCCAACCTCAGTCTGAGAGTCTTCACCGCGCGCTCTACTT 423
Db  362 CGACTTTCTGCGAGGAAACCCCAACCTCAGTCTGAGAGTCTTCACCGCGCGCTCTACTT 421
Qy  424 CTGTGAGGACCGCAAGGCTGAGCCGAGGCGGTGCGGCGCTGACCGCGCGGGGTGCA 483
Db  422 CTGTGAGGACCGCAAGGCTGAGCCGAGGCGGTGCGGCGCTGACCGCGCGGGGTGCA 481
Qy  484 AATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCA 543
Db  482 AATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCA 541
```

```
RESULT 12
AW504807
LOCUS      AW504807      570 bp      mRNA      linear      EST 02-MAR-2000
DEFINITION UI-HF-BN0-aln-f-05-0-UI.r1 NIH_MGC_50 Homo sapiens cDNA clone
IMAGE:3080216 5', mRNA sequence.
ACCESSION  AW504807
VERSION     AW504807.1  GI:7142474
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 570)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Eco RI site shown at the beginning of the sequence.
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: M.B. Soares Lab
            cDNA Library Arrayed by: M.B. Soares Lab
            DNA Sequencing by: M.B. Soares Lab
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
```

www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.
Location/Qualifiers
source
1..570

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3080216"
/clone_lib="NIH_MGC_50"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTI)"
/note="Vector: pT73-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(3.5-4.4kb). Directionally cloned. Cells provided by
Louis M. Staudt, Ph.D. Library preparation by Maria de
Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT 178 a 106 c 120 g 165 t 1 others
ORIGIN

```
    Query Match      18.6%; Score 523; DB 10; Length 570;
    Best Local Similarity 100.0%; Pred. No. 0;
    Matches 523; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy  1111 TTTCCCTTGACGTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCATGATCATAGG 1170
Db  8 TTTCCCTTGACGTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCATGATCATAGG 67
Qy  1171 ACCTCCTAATGAGAGTATCTGGGTGATTTGTGACCCCAACACCATCTCTCCAAAGCATTAAT 1230
Db  68 ACCTCCTAATGAGAGTATCTGGGTGATTTGTGACCCCAACACCATCTCTCCAAAGCATTAAT 127
Qy  1231 ATCCCAATCATGCGCTGTATGTTTAAATCAGCAGAAAGCATGTTTATGTTTCTACAAAAG 1290
Db  128 ATCCCAATCATGCGCTGTATGTTTAAATCAGCAGAAAGCATGTTTATGTTTCTACAAAAG 187
Qy  1291 AAGATTGTTATGGTGGGATGGAGGTATAGACCATGATGTCACCTTCACGCTACTTTT 1350
Db  188 AAGATTGTTATGGTGGGATGGAGGTATAGACCATGATGTCACCTTCACGCTACTTTT 247
Qy  1351 AATAAGGATCTTAAATGGCGAGGAGACTCTGTAACAGACACACCTATATATGGTTCA 1410
Db  248 AATAAGGATCTTAAATGGCGAGGAGACTCTGTAACAGACACACCTATATATGGTTCA 307
Qy  1411 TGCTGGAAGTAGCAAAATCTTCTGGAACGCAAACTCTTTAAGGAAGTCTCTAAATTTAGA 1470
Db  308 TGCTGGAAGTAGCAAAATCTTCTGGAACGCAAACTCTTTAAGGAAGTCTCTAAATTTAGA 367
Qy  1471 AACACCCCAAACTTCACATATCATATATTAGCAACAAATTTGAAGGAAGTTGCTTGAATG 1530
Db  368 AACACCCCAAACTTCACATATCATATATTAGCAACAAATTTGAAGGAAGTTGCTTGAATG 427
Qy  1531 TTGGGGAGAGGAAATCTATTGGGCTCTCTGGGTCTCTTCATCTCAGAAATGCCAATCAG 1590
Db  428 TTGGGGAGAGGAAATCTATTGGGCTCTCTGGGTCTCTTCATCTCAGAAATGCCAATCAG 487
Qy  1591 GTCAAGTTTGTGCTACATTTTGTATGTTGTGATGCTTCTTCCCA 1633
Db  488 GTCAAGTTTGTGCTACATTTTGTATGTTGTGATGCTTCTTCCCA 530
```

```
RESULT 13
BG757392
LOCUS      BG757392      693 bp      mRNA      linear      EST 15-MAY-2001
DEFINITION mRNA sequence.
ACCESSION  BG757392
VERSION     BG757392.1  GI:14068045
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE 1 (bases 1 to 693)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1694 row: k column: 05
High quality sequence stop: 693.
Location/Qualifiers
1..693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4851580"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 157 a 188 c 178 g 170 t
ORIGIN
Query Match 17.6%; Score 497; DB 12; Length 693;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 667; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 AGAAGACCATCATTAATGAAGTGAGATTTTCTGCCTGAGACTTGACGGAGGCAAGA 60
Db |
QY 2 AGAAGACCATCATTAATGAAGTGAGATTTTCTGCCTGAGACTTGACGGAGGCAAGA 61
Db |
QY 61 AGACACTCTGGACACCATATGGACAGCCTCTTGATGAACCGAGGAGAGTTCTTTTACCA 120
Db |
QY 62 AGACACTCTGGACACCATATGGACAGCCTCTTGATGAACCGAGGAGAGTTCTTTTACCA 121
Db |
QY 121 ATTCAAAATGTCCTGGCTAAGGTCGGCGTGAGACCTACCTGTGCTAGTAGTAA 180
Db |
QY 122 ATTCAAAATGTCCTGGCTAAGGTCGGCGTGAGACCTACCTGTGCTAGTAGTAA 181
Db |
QY 181 -GAGCGGTGACAGTGCTACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAAGC 239
Db |
QY 182 CGAGCGGTGACAGTGCTACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAAGC 241
Db |
QY 240 GTGCCACGTGAATGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCC 299
Db |
QY 242 GCTGCCACGTGAATGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCC 301
Db |
QY 300 GCTGCTACCGGCTACCTGCTTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATG 359
Db |
QY 302 GCTGCTACCGGCTACCTGCTTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATG 361
Db |
QY 360 TGGCCGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCT 419
Db |
QY 362 TGGCCGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCT 421
Db |
QY 420 ACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCGGTGTCACCGCGCGGG 479
Db |
QY 422 ACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCGGTGTCACCGCGCGGG 481
Db |
QY 480 TGCAAATACCATCATGACCTTCAAGAGATATTTTACTGCTGGGAATACTTTTGTAGAAA 539
Db |

Db 482 TGCAAAATAGCCATCATGACCTTCAAGAGATATATCTTTACTGCTGGAAATCTTTTGTAGAAA 541
QY |
Db 540 ACCATGAAGAACTTTTCAAGCCTCGGAGGCTCATGAAATTCAGTTCTGCTCTCCA 599
QY |
Db 542 ACCATGAAGAACTTTTCAAGCCTCGGAGGCTCATGAAATTCAGTTCTGCTCTCCA 601
QY |
Db 600 GACACCTTCGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTTACGACGACGCAATTC 659
QY |
Db 602 GACACCTTCGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTTACGACGACGCAATTC 661
QY |
Db 660 GTACTTTGG 668
QY |
Db 662 GTACTTTGG 670
QY |
RESULT 14
LOCUS BG686876 889 bp mRNA linear EST 01-MAY-2001
DEFINITION 602650861F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4763247 5',
mRNA sequence.
ACCESSION BG686876
VERSION BG686876.1 GI:13918273
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1618 row: j column: 16
High quality sequence start: 6
High quality sequence stop: 727.
Location/Qualifiers
1..889
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4763247"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 205 a 223 c 257 g 203 t
ORIGIN
Query Match 16.2%; Score 456; DB 12; Length 889;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 51 GGAGCAAGAAGACACTCTGGACACCATATGACAGCCTCTTGATGAACCGAGGAAGT 110
Db |
QY 9 GGAGCAAGAAGACACTCTGGACACCATATGACAGCCTCTTGATGAACCGAGGAAGT 68
Db |
QY 111 TTCTTTACCAATTCAAAATGTCCGCTGGGCTAAGGTCGGGTCGAGACCTACCTGTGCT 170
Db |

```
Db 69 TTCTTTACCAATTCAAAATGTCCGCTGGGCTAAGGGTCGGCTGAGACCTACCTGTGCT 128
Qy 171 ACGTAGTGAAGGGGTGACAGTGTACATCCTTTTTCACGTGACTTTGGTTATCTTCGCA 230
Db 129 ACGTAGTGAAGGGGTGACAGTGTACATCCTTTTTCACGTGACTTTGGTTATCTTCGCA 188
Qy 231 ATAAGAAGGGGTGCCACGTGGAAATTGCTTCTCCCTCGCTACATCTCGGACTGGGACCTAG 290
Db 189 ATAAGAAGGGGTGCCACGTGGAAATTGCTTCTCCCTCGCTACATCTCGGACTGGGACCTAG 248
Qy 291 ACCCTGGCGGTGCTACCGGTGACCTGTTTCACCTCTCTGGAGCCCTGCTGACTGTG 350
Db 249 ACCCTGGCGGTGCTACCGGTGACCTGTTTCACCTCTCTGGAGCCCTGCTGACTGTG 308
Qy 351 CCCGACATGTGCCGACATTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCG 410
Db 309 CCCGACATGTGCCGACATTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCG 368
Qy 411 CGCGCTCTACTTCTGTGAGGACCCCAAGGCTGAGCCGAGGGCTGCGGGCGGTGCACC 470
Db 369 CGCGCTCTACTTCTGTGAGGACCCCAAGGCTGAGCCGAGGGCTGCGGGCGGTGCACC 428
Qy 471 GCGCGGGGTGCAATAGCCATCATGACCTTCARAG 506
Db 429 GCGCGGGGTGCAATAGCCATCATGACCTTCARAG 464
```

```
RESULT 15
AW978582/c AW978582 604 bp mRNA linear EST 02-JUN-2000
LOCUS EST390591 MAGG resequences, MAGP Homo sapiens cDNA, mRNA sequence.
ACCESSION AW978582
VERSION AW978582.1 GI:8169850
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 604)
AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt
I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and
Quackenbush, J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 394
Seq primer: Forward.
FEATURES
location/Qualifiers
source
1..604
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGP"
/note="Vector: pBluescriptSKm"
BASE COUNT 190 a 121 c 109 g 184 t
ORIGIN
Query Match 15.9%; Score 448; DB 10; Length 604;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 448; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1197 TTGTGACCCCAACCATCTCTCCAAGCATTTAATCAATCATCGCTGTATGTTTAA 1256
Db 459 TTGTGACCCCAACCATCTCTCCAAGCATTTAATCAATCATCGCTGTATGTTTAA 400
Qy 1257 TCAGCAGAGCATGTTTTATCTTTGTACAAAGAGATTGTTATGGGTGGGATGGAGG 1316
Db 399 TCAGCAGAGCATGTTTTATCTTTGTACAAAGAGATTGTTATGGGTGGGATGGAGG 340
```

```
Qy 1317 TATAGACCATGATGGTCACCTTCAAGCTACTTTAATAAGGATCTTAAAAATGGCAGGA 1376
Db 339 TATAGACCATGATGGTCACCTTCAAGCTACTTTAATAAGGATCTTAAAAATGGCAGGA 280
Qy 1377 GGACTGTGAACAAGACACCCCTAATAATGGGTTGATGTCTGAAGTAGCAAACTTCTTGGAA 1436
Db 279 GGACTGTGAACAAGACACCCCTAATAATGGGTTGATGTCTGAAGTAGCAAACTTCTTGGAA 220
Qy 1437 ACGCAAACTCTTTTAAGGAAGTCCCTAATTTAGAAACACCCCAAACTTCACATATCATA 1496
Db 219 ACGCAAACTCTTTTAAGGAAGTCCCTAATTTAGAAACACCCCAAACTTCACATATCATA 160
Qy 1497 ATTACCAACAATTTGGAAGGAAGTTGCTTGAATGTTGGGAGAGGAAATCTATTGGGCTC 1556
Db 159 ATTACCAACAATTTGGAAGGAAGTTGCTTGAATGTTGGGAGAGGAAATCTATTGGGCTC 100
Qy 1557 TCGTGGGTCTCTTATCTCAGAAATGCCAATCAGGTCAAGGTTTCTACATTTGTATGT 1616
Db 99 TCGTGGGTCTCTTATCTCAGAAATGCCAATCAGGTCAAGGTTTCTACATTTGTATGT 40
Qy 1617 GTGTGATGCTTCTCCCAAGGTATATTA 1644
Db 39 GTGTGATGCTTCTCCCAAGGTATATTA 12
```

```
RESULT 16
BF975166 942 bp mRNA linear EST 22-JAN-2001
LOCUS 602244657F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:433639 5',
DEFINITION mRNA sequence.
ACCESSION BF975166
VERSION BF975166.1 GI:12342381
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 942)
AUTHORS NIH-MGC http://mgi.nhl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1207 row: a column: 16
High quality sequence stop: 707.
location/Qualifiers
source
1..942
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:433639"
/clone_lib="NIH_MGC_48"
/tissue_type="Primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
```

```
BASE COUNT 249 a 209 c 251 g 233 t
ORIGIN
```

Query Match 15.9%; Score 448; DB 12; Length 942;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 498; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 122 TTCAAAATGTCGCTGGCTAAGGCTGAGAGCTACCTGTGTACGTAGTGAAG 181
 Db 121 TTCAAAATGTCGCTGGCTAAGGCTGAGAGCTACCTGTGTACGTAGTGAAG 180

QY 182 AGGCGTGACAGTGCTACATCCTTTTCACTGGAGTTTGTATCTTCGCAATAAGAACGCG 241
 Db 181 AGGCGTGACAGTGCTACATCCTTTTCACTGGAGTTTGTATCTTCGCAATAAGAACGCG 240

QY 242 TGCCACGTGGAAATGCTCTTCCTCGCTACATCTCGGACTGGAGCTAGACCTGGCGCG 301
 Db 241 TGCCACGTGGAAATGCTCTTCCTCGCTACATCTCGGACTGGAGCTAGACCTGGCGCG 300

QY 302 TGCTACCGGTCACCTGTTACCTCTCGGAGCCCTGTAGGACTGTCCGGACATGTG 361
 Db 301 TGCTACCGGTCACCTGTTACCTCTCGGAGCCCTGTAGGACTGTCCGGACATGTG 360

QY 362 GCGGACTTTCTCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTAC 421
 Db 361 GCGGACTTTCTCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTAC 420

QY 422 TTCTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGGCTGCACCGCGCGGGTG 481
 Db 421 TTCTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGGCTGCACCGCGCGGGTG 480

QY 482 CAATAGGACATCATGACCTTCAAGATATTTTACTGCTGGAATCTTTGTAGAAAAC 541
 Db 481 CAATAGGACATCATGACCTTCAAGATATTTTACTGCTGGAATCTTTGTAGAAAAC 540

QY 542 CATGAAAGAACTTTCAAGCCCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGA 601
 Db 541 CATGAAAGAACTTTCAAGCCCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGA 600

QY 602 CAGCTTCGGCGCATCCTTT 620
 Db 601 CAGCTTCGGCGCATCCTTT 619

RESULT 17
 AIO16902/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 442)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 426 Std Error: 0.00
 Seq primer: -40m13 fwd. Et from Amersham
 High quality sequence stop: 433.
 Location/Qualifiers
 1. .442
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1627877"
 /clone_lib="Soares_NFL_T_GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBHL19W, testis NHT, and B-cell
 NCI-CGAP-GCB1) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo.

BASE COUNT 166 a 66 c 69 g 141 t
 ORIGIN

Query Match 15.7%; Score 442; DB 9; Length 442;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2211 AATGTAAGGAGCAATAAGGGATCCCTATTTGTTCTCTTTTGGTGTCTATTTGTCCTTAAC 2270
 Db 442 AATGTAAGGAGCAATAAGGGATCCCTATTTGTTCTCTTTTGGTGTCTATTTGTCCTTAAC 383

QY 2271 AACTGTCTTTGACAGTGAGAAAATATTCAGAAATACCATATCCCTGTGCGGCTTATTACC 2330
 Db 382 AACTGTCTTTGACAGTGAGAAAATATTCAGAAATACCATATCCCTGTGCGGCTTATTACC 323

QY 2331 TAGCAACCTTGCATGAAGATGAGCAGATCCACAGAGAAACTTCAATGCCACAACCTGTCT 2390
 Db 332 TAGCAACCTTGCATGAAGATGAGCAGATCCACAGAGAAACTTCAATGCCACAACCTGTCT 263

QY 2391 TATTTAATCTTATTTGTACATAAGTTTGTAAAAGAGTTTAAAATTTGTACTTTCATGTATT 2450
 Db 262 TATTTAATCTTATTTGTACATAAGTTTGTAAAAGAGTTTAAAATTTGTACTTTCATGTATT 203

QY 2451 CATTTATATTTATATTTTGGCTCTAAGATTTTATTAAACATGATTTCCCTTTCT 2510
 Db 202 CATTTATATTTATATTTTGGCTCTAAGATTTTATTAAACATGATTTCCCTTTCT 143

QY 2511 GATATATTGAAATGGAGTCTCAAGCTTCAATAATTTATACTTTAGAAATGATTTCTAAT 2570
 Db 142 GATATATTGAAATGGAGTCTCAAGCTTCAATAATTTATACTTTAGAAATGATTTCTAAT 83

QY 2571 AACACGATGTAATTTAATGTAACATTCAGTAATGTTGTCAGGAAGCCATTTCTCTTTGATT 2630
 Db 82 AACACGATGTAATTTAATGTAACATTCAGTAATGTTGTCAGGAAGCCATTTCTCTTTGATT 23

QY 2631 TTAGTAAACTTTTATGACAGCA 2652
 Db 22 TTAGTAAACTTTTATGACAGCA 1

RESULT 18
 BG170756
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 511)
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.

VERSION AA831307.1 GI:2904406
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 353)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-femail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 304.
FEATURES
 Location/Qualifiers
 source
 1..353
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1355026"
 /clone_lib="NCI-CGAP-GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD+), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTCACCAATCTGAAGTGGGAGCGCGCTCATTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
 96 a 81 c 63 g 113 t
BASE COUNT
ORIGIN
 Query Match 12.5%; Score 353; DB 9; Length 353;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1385 AACAGACACCCCTAATAATGGGTGATGCTGAAGTAGCAAACTCTTGGAAACGCAAC 1444
 Db 353 AACAGACACCCCTAATAATGGGTGATGCTGAAGTAGCAAACTCTTGGAAACGCAAC 294
 QY 1445 TCCTTTAAGGAAGTCCTTAATTTAGAAACACCCACAACTTCACATATCATATAATTAGCAA 1504
 Db 293 TCCTTTAAGGAAGTCCTTAATTTAGAAACACCCACAACTTCACATATCATATAATTAGCAA 234
 QY 1505 ACAATTGGAAGAGTTGCTTGAATGTGGGAGAGGAAATCTATTGGCTCTCGTGGGT 1564
 Db 233 ACAATTGGAAGAGTTGCTTGAATGTGGGAGAGGAAATCTATTGGCTCTCGTGGGT 174
 QY 1565 CTCTTCATCTCAGAAATGCCAATCAGGTCAAGTTTGGTACATTTTGTATGTGTGTGATG 1624
 Db 173 CTCTTCATCTCAGAAATGCCAATCAGGTCAAGTTTGGTACATTTTGTATGTGTGTGATG 114
 QY 1625 CTCTTCAGGAAGTATATTAACATATATAAGAGAGTTGTGACAAACAGAAATGATAAGCT 1684
 Db 113 CTCTTCAGGAAGTATATTAACATATATAAGAGAGTTGTGACAAACAGAAATGATAAGCT 54
 QY 1685 GCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTGGGAGGTTGAGGAGGA 1737

Db 53 GCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTGGGAGGTTGAGGAGGA 1
 RESULT 21
 AW135547/c
 LOCUS
 DEFINITION
 UI-H-BII-acq-e-01-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone
 IMAGE:2714208 3', mRNA sequence.
 AW135547
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 476)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-femail.nih.gov
 Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: M13 Forward
 POLYA=No.
FEATURES
 Location/Qualifiers
 source
 1..476
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2714208"
 /clone_lib="NCI_CGAP_Sub3"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI; The NCI_CGAP_Sub3 library is a subtracted library derived from the NCI_CGAP_Sub1 library, which is a subtracted library derived from BI. BI constitutes a mixture of 21 normalized or subtracted NCI_CGAP libraries: NCI_CGAP_Co4, NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_Co10, NCI_CGAP_Co16, NCI_CGAP_Kid5, NCI_CGAP_Kid12, NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2, NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_CLL1, NCI_CGAP_Lei2, NCI_CGAP_Brn23, NCI_CGAP_Lu5, NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6, NCI_CGAP_Brn25. These 21 libraries were pooled and a single-stranded DNA preparation of the resulting mixture was used as a tracer in a subtractive hybridization with a driver whose composition is detailed below:
 NCI_CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,
 3798-3803 (IMAGE CloneIDs 132376-132391), 1456008-1456775
 1500552-1502855); NCI_CGAP_Kid5 pool 1 LLAM 3338-3342
 3722-3725, 3776-3778 (IMAGE CloneIDs 1323912-1325831,
 1471368-1472903, 1492104-1493255); NCI_CGAP_Lu5 pool 1,
 LLAM 3575-3582, 3851-3854 (IMAGE CloneIDs 1414920-1417991,
 1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM 3164-3167,
 3716-3720, 3733-3735 (IMAGE CloneIDs 1257096-1258631,
 1469084-1470983, 1475592-1476743); NCI_CGAP_Pr22 pool 1
 LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE CloneIDs
 985608-986759, 1101192-1101959, 1217928-1220615);
 NCI_CGAP_Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE
 CloneIDs 1057416-1061255, 1144584-1145351). Subtraction
 was performed as previously described [Bonaldo, Lennon &
 Soares (1996)]: Normalization and Subtraction: Two
 Approaches to Facilitate Gene Discovery. Genome Research
 6, 791-806.
 TAG_LIB=NCI_CGAP_Lu5
 TAG_TISSUE=lung
 TAG_SEQ=CAAC

BASE COUNT		145 a	102 c	86 g	143 t
ORIGIN					
Query Match		12.4%; Score 349; DB 10; Length 476;			
Best Local Similarity		99.8%; Pred. No. 0;			
Matches		399; Conservative	0; Mismatches	1; Indels	0; Gaps
QY	832	TATGCTCTGAAATACAGAGGACAGCTCTGGCCAGGAGCTGCAATTGGTGCA			
DB	476	TATGCTCTGAAATACAGAGGACAGCTCTGGCCAGGAGCTGCAATTGGTGCA			
QY	892	GTTTGAATGCAACATGTCCTCTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATC			
DB	416	GTTTGAATGCAACATGTCCTCTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATC			
QY	952	CTAAAGTGTCAAGCTTTTCTATGACTTTTAGTAGATGAGAGCAGAGGTAGATCCTA			
DB	356	CTAAAGTGTCAAGCTTTTCTATGACTTTTAGTAGATGAGAGCAGAGGTAGATCCTA			
QY	1012	AAAAGCATGGTGAGAGATCAAAATGTTTATATCAACATCCCTTTATTTGATTCATT			
DB	296	AAAAGCATGGTGAGAGATCAAAATGTTTATATCAACATCCCTTTATTTGATTCATT			
QY	1072	TGAGTTAAACAGTGGTGTAGTAGATGATTTTCTATTTTCCCTTGAGCTTTACTTTC			
DB	236	TGAGTTAAACAGTGGTGTAGTAGATGATTTTCTATTTTCCCTTGAGCTTTACTTTC			
QY	1132	AAGTACACAAACTCTCCATCAGCCATGATCTATAGGACCTCCTAAATCAGAGTATCTG			
DB	176	AAGTACACAAACTCTCCATCAGCCATGATCTATAGGACCTCCTAAATCAGAGTATCTG			
QY	1192	GGTGATTGTGACCCCAAAACCATCTCTCCAAAGCATTAATA			
DB	116	GGTGATTGTGACCCCAAAACCATCTCTCCAAAGCATTAATA			
RESULT 22					
AW452648/c					
LOCUS		AW452648			
DEFINITION		UT-H-B13-alu-f-11-0-UT.sl NCI_CGAP_Sub5 Homo sapiens cDNA clone			
ACCESSION		IMAGE:3068708 3', mRNA sequence.			
VERSION		AW452648.1 GI:6993424			
KEYWORDS		EST.			
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS		1 (bases 1 to 464)			
TITLE		NCI-Cgap http://www.ncbi.nlm.nih.gov/ncicgap.			
JOURNAL		National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index			
COMMENT		Unpublished (1997)			
		Contact: Robert Strausberg, Ph.D.			
		Email: cgaps-r@mail.nih.gov			
		The sequence contained an oligo-dT track that was present in the			
		oligonucleotide that was used to prime the synthesis of first			
		strand cDNA and therefore this may represent a bonafide poly A			
		tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:			
		NCI-Cgap clone distribution information can be found through the			
		I.M.A.G.E. Consortium/LLNL at:			
		www.bio.lnl.gov/dbp/image/image.html The following repetitive			
		elements were found in this cDNA sequence: 4-123, >ALU			
		Seq primer: M13 Forward			
		POLYA=Yes.			
FEATURES		Location/Qualifiers			
source		1..464			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
		/clone="IMAGE:3068708"			
		/clone_lib="NCI_CGAP_Sub5"			
		/lab_host="DH10B (Life Technologies)"			
		/note="Vector: pT73D-Pac (Pharmacia) with a modified			

polylinker; Site_1: Not 1; Site_2: Eco RI; NCI_CGAP_Sub5		is a subtracted library derived from NCI_CGAP_Sub4. The			
NCI_CGAP_Sub5 library had 3 million recombinants. A		single-stranded DNA preparation of NCI_CGAP_Sub4 was used			
as a tracer in a subtractive hybridization with a driver		comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 IJAM			
3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)		132376-132391, 1456008-1456775, 1500552-1502855;			
NCI_CGAP_Kid5 pool 1 LAM 3338-3342, 3722-3725, 3776-3778		(IMAGE Clones)			
1492104-1493255; NCI_CGAP_Lu5 pool 1 LAM 3575-3582,		3851-3854 (IMAGE Clones)			
); NCI_CGAP_GC4 pool 1 LAM 3164-3167, 3716-3720,		3733-3735 (IMAGE Clones)			
1475592-1476743; NCI_CGAP_P+22 pool 1 LAM 2457-2459,		2758-2759, 3062-3068 (IMAGE Clones)			
1101192-1101959, 1217928-1220615; NCI_CGAP_Co10 pool 1		LAM 2644-2653, 2871-2872 (IMAGE Clones)			
, 1144584-1145351). (10% of the driver population), plus a		pool of 3,840 arrayed clones from NCI_CGAP_Sub1 (IMAGE			
Clones)		2710536-2712455) (10% of the driver population			
), plus a pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE		Clones)			
2712456-2723591) (10% of the driver population),		plus a pool of 5,472 clones from NCI_CGAP_Sub4 (IMAGE			
Clones)		2723592-2728969) (70% of the driver population).			
Subtraction was performed as previously described (Bonaldi		Lennon & Soares (1996): Normalization and Subtraction:			
Two Approaches To Facilitate Gene Discovery. Genome		Research 6, 791-806.			
TAG_LIB=NCI_CGAP_Lym2		TAG_TISSUE=lymph node			
TAG_SEQ=AAATG"					
BASE COUNT		114 a	111 c	84 g	155 t
ORIGIN					
Query Match		12.3%; Score 346; DB 10; Length 464;			
Best Local Similarity		100.0%; Pred. No. 0;			
Matches		346; Conservative	0; Mismatches	0; Indels	0; Gaps
QY	1361	CTTAAATGGCAGGAGGACTGTGAACAACACACCTTAATTAATGGTGTGATCTGAAGT			
DB	454	CTTAAATGGCAGGAGGACTGTGAACAACACACCTTAATTAATGGTGTGATCTGAAGT			
QY	1421	AGCAATCTTCTTGGAAACGCAAACTCTTTTAAGGAAGTCCCTTAATTTAGAAACACCCACA			
DB	394	AGCAATCTTCTTGGAAACGCAAACTCTTTTAAGGAAGTCCCTTAATTTAGAAACACCCACA			
QY	1481	AAGTTCACATATCATATATTAGCAACAATTTGGAAGGAAGTTGCTTGAATGTTGGGAGAG			
DB	334	AAGTTCACATATCATATATTAGCAACAATTTGGAAGGAAGTTGCTTGAATGTTGGGAGAG			
QY	1541	GAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGTCAAGTTT			
DB	274	GAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGTCAAGTTT			
QY	1601	GCTACATTTTGTATGCTGTGATGCTTCTCCCAAAGGTATATTAACTATATAGAGAGTT			
DB	214	GCTACATTTTGTATGCTGTGATGCTTCTCCCAAAGGTATATTAACTATATAGAGAGTT			
QY	1661	GTGCAAAACAGATGATAAAGCTGCCAACCGTGGCACACGCTCAT			
DB	154	GTGCAAAACAGATGATAAAGCTGCCAACCGTGGCACACGCTCAT			
RESULT 23					
AI468242/c					
LOCUS		AI468242			
DEFINITION		tg35905.x1 Soares_NFL_T_GBC-S1 Homo sapiens cDNA clone			
		IMAGE:2110808 3' similar to contains Alu repetitive element;; mRNA			
		sequence.			
ACCESSION		AI468242			
VERSION		AI468242.1 GI:4330332			


```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1. (bases 1 to 321)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1095 Std Error: 0.00
Seq primer: -400p from Gibco
High quality sequence stop: 304.
FEATURES
source
Location/Qualifiers
1. .321
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2110808"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCI-CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 84 a 81 c 56 g 100 t
ORIGIN
Query Match 11.4%; Score 321; DB 9; Length 321;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1488 CATATCATATAATAGCAACAATTTGGAAGGAGTTGCTTGAATCTTGGGAGAGGAAATC 1547
Db 321 CATATCATATAATAGCAACAATTTGGAAGGAGTTGCTTGAATCTTGGGAGAGGAAATC 262
Qy 1548 TATTGGCTCTCGTGGGTCTTTCATCTCAGAAATGCCAATCAGGTCAAGTTTGTCTACAT 1607
Db 261 TATTGGCTCTCGTGGGTCTTTCATCTCAGAAATGCCAATCAGGTCAAGTTTGTCTACAT 202
Qy 1608 TTTGTATGTGTGTGTCTCTCCAAAGGTATATTAACTATATAAGAGAGTTGTGACAA 1667
Db 201 TTTGTATGTGTGTGTCTCTCCAAAGGTATATTAACTATATAAGAGAGTTGTGACAA 142
Qy 1668 AACAGAATGATAAGCTGGCAACCGTGGCACACGCTCATAGTTCTAGCTTCITGGGAGGT 1727
Db 141 AACAGAATGATAAGCTGGCAACCGTGGCACACGCTCATAGTTCTAGCTTCITGGGAGGT 82
Qy 1728 TGAGGAGGAGGATGCTTGAACACACAGGTGTTCAAGCCAGCCTGGGCAACATAACAAGA 1787
Db 81 TGAGGAGGAGGATGCTTGAACACACAGGTGTTCAAGCCAGCCTGGGCAACATAACAAGA 22
Qy 1788 TCCTGTCTCTCAAAAAA 1808
Db 21 TCCTGTCTCTCAAAAAA 1
RESULT 24
AW768364 413 bp mRNA linear EST 04-MAY-2000
LOCUS hK73004.x1 NCI-CGAP_Lym12 Homo sapiens cDNA clone IMAGE:3002311 3',
DEFINITION mRNA sequence.

```

```

ACCESSION AW768364
VERSION AW768364.1 GI:7700383
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1. (bases 1 to 413)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Life Technologies catalog #: 11547-015
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -400p from Gibco
High quality sequence stop: 411.
FEATURES
source
Location/Qualifiers
1. .413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3002311"
/clone_lib="NCI-CGAP_Lym12"
/tissue_type="lymphoma, follicular mixed small and large
cell"
/lab_host="DH10B"
/notes="Organ: lymph node; Vector: pCMV-SPORT6; Site_1:
SalI; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dt. Average insert size 1.25 kb. Life Technologies
catalog #: 11547-015"
BASE COUNT 127 a 72 c 94 g 118 t 2 others
ORIGIN
Query Match 11.1%; Score 312; DB 10; Length 413;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1282 GTCAACAAGAAGATTGTTATGGTGGGATGGAGGTATAGACCATGATGGTCACTTCA 1341
Db 43 GTCAACAAGAAGATTGTTATGGTGGGATGGAGGTATAGACCATGATGGTCACTTCA 102
Qy 1342 AGCTACTTTAATAAAGGATCTTAAATGGCAGGAGGACTGTGAACAAGACACCCCTAATA 1401
Db 103 AGCTACTTTAATAAAGGATCTTAAATGGCAGGAGGACTGTGAACAAGACACCCCTAATA 162
Qy 1402 ATGGGTTGATGTCTGAAGTAGCAAACTCTTCTGGAAACGCAAACTCTTTTAAAGGAAGTCCC 1461
Db 163 ATGGGTTGATGTCTGAAGTAGCAAACTCTTCTGGAAACGCAAACTCTTTTAAAGGAAGTCCC 222
Qy 1462 TAATTTAGAAACACCCACAACTTCACATATCATATAATTTAGCAACAATTTGGAAGGAGTT 1521
Db 223 TAATTTAGAAACACCCACAACTTCACATATCATATAATTTAGCAACAATTTGGAAGGAGTT 282
Qy 1522 GCTTTGAATGTTGGGAGAGGAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCAGAAAT 1581
Db 283 GCTTTGAATGTTGGGAGAGGAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCAGAAAT 342
Qy 1582 GCCAATCAGGTC 1593
Db 343 GCCAATCAGGTC 354
RESULT 25
AI249134/c 363 bp mRNA linear EST 01-DEC-1998
LOCUS qh79c07.x1 Soares_fetal_liver_spleen_INFLS_S1 Homo sapiens cDNA
DEFINITION clone IMAGE:1850892 3' similar to contains Alu repetitive element
;contains element L1 repetitive element ;, mRNA sequence.
ACCESSION AI249134

```

```
VERSION      AI249134.1  GI:3844531
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 363)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             This clone is available royalty-free through LLNL; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Insert length: 1487, Std Error: 0.00
             Seq primer: -400P from Gibco
             High quality sequence stop: 360.
FEATURES     Location/Qualifiers
             source
               1..363
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:1850892"
               /clone_lib="Soares_fetal_liver_spleen_1NPLS_S1"
               /sex="male"
               /dev_stage="20 week-post conception fetus"
               /lab_host="DH10B (ampicillin resistant)"
               /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
               with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
               This is a subtracted version of the original Soares fetal
               liver spleen INFLS library. 1st strand cDNA was primed
               with a Pac I - oligo(dT) primer [5',
               AACTGGAAGAATTAATTAAGATCTTTTGTGTTT 3'],
               double-stranded cDNA was ligated to Eco RI adaptors
               (Pharmacia), digested with Pac I and cloned into the Pac I
               and Eco RI sites of the modified pT7T3 vector. Library
               went through one round of normalization. Library
               constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT   95 a 85 c 68 g 115 t
ORIGIN
Query Match 10.9%; Score 308; DB 9; Length 363;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 308; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1500 AGCAACAAATTGGAAGGAGTTGCTTGAATGTTGGGAGAGGAGAAATCTATTGGCTCTCG 1559
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 308 AGCAACAAATTGGAAGGAGTTGCTTGAATGTTGGGAGAGGAGAAATCTATTGGCTCTCG 249
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1560 TGGGTCTCTTCATCTCGAAATGCCAATCAGGTCAAGGTTTGGCTACATTTTGTATGTG 1619
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 248 TGGGTCTCTTCATCTCGAAATGCCAATCAGGTCAAGGTTTGGCTACATTTTGTATGTG 189
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1620 TGATGCTTCTCCCAAGGATATTAATCTATATAGAGAGTTGTGACAAAACAGATGATA 1679
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 188 TGATGCTTCTCCCAAGGATATTAATCTATATAGAGAGTTGTGACAAAACAGATGATA 129
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1680 AAGCTGGCAACCGTGGCACACGCTCATAGTTCTTGGGAGGTTGAGGAGGAGG 1739
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 128 AAGCTGGCAACCGTGGCACACGCTCATAGTTCTTGGGAGGTTGAGGAGGAGG 69
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1740 ATGCTTTGAACACAGGTTTCAAGGCCAGCGCTGGGCAACATAACAAAGATCCTGCTCTCA 1799
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 68 ATGCTTTGAACACAGGTTTCAAGGCCAGCGCTGGGCAACATAACAAAGATCCTGCTCTCA 9
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1800 AAAAAAA 1807
Db |||||||
Qy 8 AAAAAAA 1
Db |||
RESULT 26
BF664355 LOCUS 956 bp mRNA linear EST 21-DEC-2000
```

```
Db      522 CAGGCCATGATCTATAGGACCTCCCTAAATGAGAGTATCTGGGTGATTGTGACCCCAAC 579
|||||
|||||

RESULT 27
AW263139/c
LOCUS      422 bp      mRNA      linear      EST 28-DEC-1999
DEFINITION      xn78d03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2700581 3', mRNA sequence.
ACCESSION      AW263139
VERSION        AW263139.1 GI:6639955
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 422)
REFERENCE      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE         Tumor Gene Index
JOURNAL       Unpublished (1997)
COMMENT       Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-remail.nih.gov
               This clone is available royalty-free through LNL : contact the
               IMAGE Consortium (info@image.llnl.gov) for further information.
               Seq primer: -40UP from Gibco
               High quality sequence stop: 419.
               Location/Qualifiers
               1..422
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:2700581"
               /clone_lib="Soares_NFL_T_GBC_S1"
               /lab_host="DH10B"
               /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
               a modified polylinker; Site_1: Not I; Site_2: Eco RI;
               Equal amounts of plasmid DNA from three normalized
               libraries (fetal lung NbHL19W, testis NHT, and B-cell
               NCI-CGAP GCBI) were mixed, and ss circles were made in
               vitro. Following HAP purification, this DNA was used as
               tracer in a subtractive hybridization reaction. The driver
               was PCR-amplified cDNAs from pools of 5,000 clones made
               from the same 3 libraries. The pools consisted of
               I.M.A.G.E. clones 297480-302087, 682632-687239,
               726408-728711, and 729096-731399. Subtraction by Bento
               Soares and M. Fatima Bonaldo."
               136 a 90 c 79 g 117 t

BASE COUNT      136 a 90 c 79 g 117 t
ORIGIN
Query Match      10.5%; Score 296; DB 10; Length 422;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 346; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      885 TGGTGCAGTTTTGAATGCAACATTTGCCCTACTGGGAATACAGAACTGCAGGACCTGG 944
Db      397 TGGTGCAGTTTTGAATGCAACATTTGCCCTACTGGGAATACAGAACTGCAGGACCTGG 338
QY      945 GAGCATCCTAAAGTGTCACAGTTTTTCTATGACTTTTATAGTGGATGAGAGCAGAGGTA 1004
Db      337 GAGCATCCTAAAGTGTCACAGTTTTTCTATGACTTTTATAGGAGGATGAGAGCAGAGGTA 278
QY      1005 GATCCTAAAAGCATGGTGAGAGGATCAAAATGTTTTTATATCAACATCCTTTATTATTGG 1064
Db      277 GATCCTAAAAGCATGGTGAGAGGATCAAAATGTTTTTATATCAACATCCTTTATTATTGG 218
QY      1065 ATTCATTTGAGTTAAACAGTGGTGTAGTGATAGATTTTCTATCTTTCCCTTGGACGTT 1124
Db      217 ATTCATTTGAGTTAAACAGTGGTGTAGTGATAGATTTTCTATCTTTCCCTTGGACGTT 158
QY      1125 TACTTTCAAGTAACACAAACTCTTCATCAGGCCATGATCTATAGGACCTCTTAATGAGA 1184
Db      157 TACTTTCAAGTAACACAAACTCTTCATCAGGCCATGATCTATAGGACCTCTTAATGAGA 98
|||||
|||||

RESULT 28
AA954956/c
LOCUS      528 bp      mRNA      linear      EST 07-JUL-1998
DEFINITION      op24a12.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1577758 3', mRNA sequence.
ACCESSION      AA954956
VERSION        AA954956.1 GI:3118651
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 528)
REFERENCE      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE         Tumor Gene Index
JOURNAL       Unpublished (1997)
COMMENT       Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-remail.nih.gov
               This clone is available royalty-free through LNL : contact the
               IMAGE Consortium (info@image.llnl.gov) for further information.
               Insert Length: 643 Std Error: 0.00
               Seq primer: -40ml3 fwd ET from Amersham
               High quality sequence stop: 443.
               Location/Qualifiers
               1..528
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:1577758"
               /clone_lib="Soares_NFL_T_GBC_S1"
               /lab_host="DH10B"
               /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
               a modified polylinker; Site_1: Not I; Site_2: Eco RI;
               Equal amounts of plasmid DNA from three normalized
               libraries (fetal lung NbHL19W, testis NHT, and B-cell
               NCI-CGAP GCBI) were mixed, and ss circles were made in
               vitro. Following HAP purification, this DNA was used as
               tracer in a subtractive hybridization reaction. The driver
               was PCR-amplified cDNAs from pools of 5,000 clones made
               from the same 3 libraries. The pools consisted of
               I.M.A.G.E. clones 297480-302087, 682632-687239,
               726408-728711, and 729096-731399. Subtraction by Bento
               Soares and M. Fatima Bonaldo."
               173 a 101 c 100 g 150 t

BASE COUNT      173 a 101 c 100 g 150 t
ORIGIN
Query Match      10.5%; Score 296; DB 9; Length 528;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 346; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      885 TGGTGCAGTTTTGAATGCAACATTTGCCCTACTGGGAATACAGAACTGCAGGACCTGG 944
Db      397 TGGTGCAGTTTTGAATGCAACATTTGCCCTACTGGGAATACAGAACTGCAGGACCTGG 338
QY      945 GAGCATCCTAAAGTGTCACAGTTTTTCTATGACTTTTATAGTGGATGAGAGCAGAGGTA 1004
Db      337 GAGCATCCTAAAGTGTCACAGTTTTTCTATGACTTTTATAGGAGGATGAGAGCAGAGGTA 278
QY      1005 GATCCTAAAAGCATGGTGAGAGGATCAAAATGTTTTTATATCAACATCCTTTATTATTGG 1064
Db      277 GATCCTAAAAGCATGGTGAGAGGATCAAAATGTTTTTATATCAACATCCTTTATTATTGG 218
QY      1065 ATTCATTTGAGTTAAACAGTGGTGTAGTGATAGATTTTCTATCTTTCCCTTGGACGTT 1124
Db      217 ATTCATTTGAGTTAAACAGTGGTGTAGTGATAGATTTTCTATCTTTCCCTTGGACGTT 158
QY      1125 TACTTTCAAGTAACACAAACTCTTCATCAGGCCATGATCTATAGGACCTCTTAATGAGA 1184
Db      1125 TACTTTCAAGTAACACAAACTCTTCATCAGGCCATGATCTATAGGACCTCTTAATGAGA 1184
|||||
|||||
```

Db 157 TACTTTCAAGTAACACAACTCTTCCATCAGGCCATGATCTATAGGAACCTCTTAATGAGA 98

QY 1185 GTATCTGGTGATGTCAGCCCAACCATCTCTCCAAAGCATTAAATA 1231

Db 97 GTATCTGGTGATGTCAGCCCAACCATCTCTCCAAAGCATTAAATA 51

RESULT 29

AA504653/c

LOCUS

DEFINITION

aa63f11.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:825645 3'

similar to contains Alu repetitive element; , mRNA sequence.

ACCESSION

AA504653

VERSION

AA504653.1 GI:2240813

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 332)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL

Tumor Gene Index

COMMENT

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., David Allman,

Ph.D., Gerald Marti, M.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40m13 fwd. Et from Amersham

High quality sequence stop: 294.

Location/Qualifiers

1. .332

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:825645"

/clone_lib="NCI_CGAP_GCB1"

/tissue_type="germinal center B cell"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was prepared from human tonsillar cells enriched for

germinal center B cells by flow sorting (CD20+, IgD+),

provided by Dr. Louis M. Staudt (NCI), Dr. David Allman

(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was

primed with a Not I - oligo(dT) primer

15'-NGTACCAATCTGAAGTGGAGCGCGCTCATTTTCTTTTCTTT-3'

1. Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

86 a 82 c 61 g 103 t

ORIGIN

Query Match 10.0%; Score 281; DB 9; Length 332;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 331; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1476 CCACAACTTCATATATATATAGCAAACTTGAAGGAGTGTGTAATGTGGG 1535

Db 332 CCACAACTTCATATATATATAGCAAACTTGAAGGAGTGTGTAATGTGGG 273

QY 1536 GAGAGGAAATCTATTGGCTCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAA 1595

Db 272 GAGAGGAAATCTATTGGCTCGTGGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAA 213

QY 1596 GGTTCGTACATATTTTGTATGTGTGATGCTTCTCCCAAGAGTATATTAACATATATAAGA 1655

Db 212 GGTTCGTACATATTTTGTATGTGTGATGCTTCTCCCAAGAGTATATTAACATATATAAGA 153

QY 1656 GAGTTGTGACAAACACAGATGATAAAGCTGCCAACCGTGGCACACGCTCATAGTTCTAGC 1715

Db 152 GAGTTGTGACAAACACAGATGATAAAGCTGCCAACCGTGGCACACGCTCATAGTTCTAGC 93

QY 1716 TCCTTGGGAGGTTGAGGAGGAGGATGCTTCAACACACAGGTTTCAAGGTCAGCCTGGGC 1775

Db 92 TCCTTGGGAGGTTGAGGAGGAGGATGCTTCAACACACAGGTTTCAAGGTCAGCCTGGGC 33

QY 1776 AACATAACAAGATCCTGCTCTCTCAAAAAAAA 1807

Db 32 AACATAACAAGATCCTGCTCTCTCAAAAAAAA 1

RESULT 30

BG398364

LOCUS

DEFINITION

602439832F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4566366 5',

mRNA sequence.

ACCESSION

BG398364

VERSION

BG398364.1 GI:13291812

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 928)

AUTHORS

NIH-MGC http://mgc.nci.nih.gov/.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incey Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM1284 row: o column: 07

High quality sequence stop: 650.

Location/Qualifiers

1. .928

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4566366"

/clone_lib="NIH_MGC_48"

/tissue_type="primary B-cells from tonsils (cell line)"

/lab_host="DH10B (phage-resistant)"

/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;

Site_2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC Library."

BASE COUNT

280 a 163 c 222 g 263 t

ORIGIN

Query Match 9.9%; Score 279; DB 12; Length 928;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 519; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 665 TTGGGACTTTGATAGCAACTTCCAGGAATGTCCACACAGCATCAATATCTCTGCTGAACA 724

Db 97 TTGGGACTTTGATAGCAACTTCCAGGAATGTCCACACAGCATCAATATCTCTGCTGAACA 156

QY 725 CAGTGGATAAAAAACAGTCCTTCAAGTCTCTCTGTTTATTCTTCAAC(TCTCACTTTC 784

|||||
Db 157 CAGTGGATAAAACAGTCTCTCAAGTCTCTCTGTTTATTTCTTCAACTCTCACTTTC 216
QY 785 TTAGAGTTTACAGAAAAATATTTATATACGACTCTTTTAAAGATCTATGCTTCAAAA 844
Db 217 TTAGAGTTTACAGAAAAATATTTATATACGACTCTTTTAAAGATCTATGCTTCAAAA 276
QY 845 TAGAGAAGAACACAGGCTGGCCAGGAGCTGCTGCAATTTGGTGCAGTTTTGAATGCAA 904
Db 277 TAGAGAAGAACACAGGCTGGCCA -GGAGTGTCTGCAATTTGGTGCAGTTTTGAATGCAA 335
QY 905 CATTGTCCTACTGGATACAGAACTGCAGGACCTGGAGCATCTTAAAGTGTCAAC 964
Db 336 CATTGTCCTACTGGGAATAACAGAACTGCAGGACCTGGAGCATCTTAAAGTGTCAAC 395
QY 965 GTTTTTCTATGACTTTTAGTAGGATGAGAGCAGAGGTAGATCCTTAAAGCATGCTGA 1024
Db 396 GTTTTTCTATGACTTTTAGTAGGATGAGAGCAGAGGTAGATCCTTAAAGCATGCTGA 455
QY 1025 GAGGATCAATG-TTTTTATATCAACATCCTTTATTATTGATTCATTTGAGTTAACAGT 1083
Db 456 GAGGATCAATGTTTTTATATCAACATCCTTTATTATTGATTCATTTGAGTTAACAGT 515
QY 1084 GGTGTTAGTAGATAGATTTTCTATTCTTTCCCTTGACGTTTACTTTTCAAGTAAACAAA 1143
Db 516 GGTGTTAGTAGATAGATTTTCTATTCTTTCCCTTGACGTTTACTTTTCAAGTAAACAAA 575
QY 1144 CTCTTCCATCAGGCCATGATCTATAGGACTCTCTAATGAGA 1184
Db 576 CTCTTCCATCAGGCCATGATCTATAGGACTCTCTAATGAGA 616

RESULT 31
AW134750
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html The following repetitive
elements were found in this cDNA sequence: 1-21,
>At_rich#Low_complexity
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source
1..454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2712644"
/clone_lib="NCI-CGAP_Sub3"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NCI-CGAP_Sub3 library is a subtracted library derived from
the NCI-CGAP_Sub1 library, which is a subtracted library

derived from BI. BI constitutes a mixture of 21
normalized or subtracted NCI-CGAP libraries: NCI-CGAP_Co4
, NCI-CGAP_Pr22, NCI-CGAP_Pr28, NCI-CGAP_Co10,
NCI-CGAP_Co16, NCI-CGAP_Kid5, NCI-CGAP_Kid12,
NCI-CGAP_Kid3, NCI-CGAP_Kid11, NCI-CGAP_Lym2,
NCI-CGAP_Brn2, NCI-CGAP_Co8, NCI-CGAP_CLL1, NCI-CGAP_Lei2,
NCI-CGAP_Brn23, NCI-CGAP_Lu5, NCI-CGAP_Lu24,
NCI-CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI-CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clones 1322376-1323911, 1456008-1456775
, 1500552-1502855); NCI-CGAP_Kid5 pool 1 LLAM 3338-3342
, 3722-3725, 3776-3778 (IMAGE Clones 1323912-1325831,
1471368-1472903, 1492104-1493255); NCI-CGAP_Lu5 pool 1,
LLAM 3575-3582, 3851-3854 (IMAGE Clones 1414920-1417991,
1520904-1522439); NCI-CGAP_GC4 pool 1 LLAM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clones 1257096-1258631,
1469064-1470983, 1475592-1476743); NCI-CGAP_Pr22 pool 1
LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clones
985608-986759, 1101192-1101959, 1217928-1220615);
NCI-CGAP_Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE
Clones 1057416-1061255, 1144584-1145351). Subtraction
was performed as previously described (Bonaldi, Lennon &
Soares (1996): Normalization and Subtraction: Two
Approaches To Facilitate Gene Discovery. Genome Research
6, 791-806.
TAG_LIB=NCI-CGAP_GC4
TAG_TISSUE=germ cell
TAG_SEQ=AAATC"

BASE COUNT 136 a 105 c 68 g 145 t
ORIGIN
Query Match 9.9%; Score 278; DB 10; Length 454;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 278; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 622 GCCCTGTATGAGTTGATGACTTACGAGCAGCATTTCTGCTGGAACAGTGGATAAAACAG 681
Db 177 GCCCTGTATGAGTTGATGACTTACGAGCAGCATTTCTGCTGGAACAGTGGATAAGCA 236
QY 682 ACTTCCAGGAATGTCACACACGATGAATATCTCTGCTGAAGACAGTGGATAAAACAG 741
Db 237 ACTTCCAGGAATGTCACACACGATGAATATCTCTGCTGAAGACAGTGGATAAAACAG 296
QY 742 TCTTCAAGTCTTCTCTGTTTTTATTTCTTCAACTCTCACTTCTTTAGAGTTTACAGAAA 801
Db 297 TCTTCAAGTCTTCTCTGTTTTTATTTCTTCAACTCTCACTTCTTTAGAGTTTACAGAAA 356
QY 802 AATATTATATACGACTCTTAAAGATCTATGCTTCAAAATAGAGGAGGACACAGG 861
Db 357 AATATTATATACGACTCTTAAAGATCTATGCTTCAAAATAGAGGAGGACACAGG 416
QY 862 TCTGGCCAGGAGCTGCTGCAATTTGGTGCAGTTTTGAA 899
Db 417 TCTGGCCAGGAGCTGCTGCAATTTGGTGCAGTTTTGAA 454

RESULT 32
BG341819
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
BG341819 939 bp mRNA linear EST 27-FEB-2001
602463552F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4576188 5',
mRNA sequence.
BG341819
BG341819.1 GI:13148257
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 939)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCMI288 row: h column: 13
High quality sequence stop: 547.
Location/Qualifiers
1..939
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4576188"
/clone_lib="NIH_MGC_48"
/tissue.type="primary B-cells from tonsils (cell line)"
/lab.host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 253 a 196 c 222 g 268 t
ORIGIN

Query Match 7.3%; Score 207; DB 12: Length 939;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 665 TTGGGACTTTGTAGCAACTCCAGGAATGTCACACACATGAATATCTCTGCTGAAGA 724
Db 98 TTGGGACTTTGTAGCAACTCCAGGAATGTCACACACATGAATATCTCTGCTGAAGA 157
QY 725 CAGTGGATAAAAACAGTCTTCAAGTCTCTCTGTTTATATCTTCAACTCTCACTTTC 784
Db 158 CAGTGGATAAAAACAGTCTTCAAGTCTCTCTGTTTATATCTTCAACTCTCACTTTC 217
QY 785 TTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTCAAAA 844
Db 218 TTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTGAAGA 277
QY 845 TAGAGAAGGAACACAGGCTGGCCAGG 871
Db 278 TAGAGAAGGAACACAGGCTGGCCAGG 304

RESULT 33
LOCUS BG170824
DEFINITION 602323763F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4426959 5', mRNA sequence.
ACCESSION BG170824 GI:12677527
VERSION BG170824.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 517)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov

Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCMI10174 row: n column: 16
High quality sequence stop: 517.
Location/Qualifiers
1..517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4426959"
/clone_lib="NIH_MGC_89"
/tissue.type="hypernephroma, cell line"
/lab.host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 kb. Library enriched for full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 181 a 94 c 118 g 124 t
ORIGIN

Query Match 7.3%; Score 205; DB 12: Length 517;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1599 TTGCTACATTTTGTATGTGTGATGCTTCTCCCAAGGATATATTAACATATATAAGAGAG 1658
Db 283 TTGCTACATTTTGTATGTGTGATGCTTCTCCCAAGGATATATTAACATATATAAGAGAG 342
QY 1659 TTGTGACAAAACAGATGATAAAGCTGCGAACGCTGGCACAGCTCATAGTTCTAGCTGC 1718
Db 343 TTGTGACAAAACAGATGATAAAGCTGCGAACGCTGGCACAGCTCATAGTTCTAGCTGC 402
QY 1719 TTGGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGTTTCAAGCCACGCTGGGCAAC 1778
Db 403 TTGGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGTTTCAAGCCACGCTGGGCAAC 462
QY 1779 ATAACAAGATCTGCTCTCTCAAAA 1803
Db 463 ATAACAAGATCTGCTCTCTCAAAA 487

RESULT 34
LOCUS AW444889
DEFINITION UI-H-B13-ajz-c-08-0-UI.s1 NCL_CGAP_Sub5 Homo sapiens cDNA clone IMAGE:2733327 3', mRNA sequence.
ACCESSION AW444889
VERSION AW444889.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 215)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 507)
AUTHORS Li.W.B., Gruber.C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Location/Qualifiers
1..507
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DG003YB14"
/clone_lib="LRI_FL011_Bc1"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/lab_host="DH10B"
/note="vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850 USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 134 a 123 c 84 g 158 t 8 others
ORIGIN
Query Match 5.4%; Score 153; DB 9; Length 507;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1412 GTCGGAAGTAGCAATCTCTGGAACGCAAACTCTTTAAGGAAGTCCCTTAATTAGAA 1471
Db 366 GTCGGAAGTAGCAATCTCTGGAACGCAAACTCTTTAAGGAAGTCCCTTAATTAGAA 307
QY 1472 ACACCCACAATTCACATATCATATTTAGCAACAATTTGAAGGAAGTTCCTTGAATGT 1531
Db 306 ACACCCACAATTCACATATCATATTTAGCAACAATTTGAAGGAAGTTCCTTGAATGT 247
QY 1532 TGGGAGAGGAGAAATCTATTGGCTCTCGTGGGT 1564
Db 246 TGGGAGAGGAGAAATCTATTGGCTCTCGTGGGT 214

RESULT 37
AA879422/c
LOCUS 268 bp mRNA linear EST 19-MAY-1998
DEFINITION oJ91c11.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1505684 3', mRNA sequence.
ACCESSION AA879422
VERSION AA879422.1 GI:2988533
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 268)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 685 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 252.

FEATURES
source
Location/Qualifiers
1..268
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1505684"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NDHL19W, testis NHT, and B-cell NCI-CGAP-GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 91 a 40 c 30 g 107 t
ORIGIN
Query Match 5.2%; Score 147; DB 9; Length 268;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 267; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 2518 TGAATGAGCTCANAAGCTTCATAAATTTATACTTTAGAAATGATTCTAATAACAACG 2577
Db 268 TGAATGAGCTCANAAGCTTCATAAATTTATACTTTAGAAATGATTCTAATAACAACG 209
QY 2578 TATGTAATTTGTAACATTGCAGTAAATGCTGCTACGAAGCCATTTCTCTTGATTCTTTAGTAA 2637
Db 208 TATGTAATTTGTAACATTGCAGTAAATGCTGCTACGAAGCCATTTCTCTTGATTCTTTAGTAA 149
QY 2638 ACTTTTA-TGACAGCAAAATTTGCTTCGCGTCACATTTCAATCAGTTAAATTAATCATATAA 2696
Db 148 ACTTTTATTTGACAGCAAAATTTGCTTCGCGTCACATTTCAATCAGTTAAATTAATCATATAA 89
QY 2697 TAAATTTTGAAGCTGTGAAGATAAAATACCAATAAATAATATAAATAATATAAATATATATATG 2756
Db 88 TAAATTTTGAAGCTGTGAAGATAAAATACCAATAAATAATATAAATAATATAAATATATATATG 29
QY 2757 AAGTTAAATAAATAATCAGTATGATGG 2784
Db 28 AAGTTAAATAAATAATCAGTATGATGG 1

RESULT 38
BF664352
LOCUS 1140 bp mRNA linear EST 21-DEC-2000
DEFINITION 602146010F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4309563 5', mRNA sequence.
ACCESSION BF664352
VERSION BF664352.1 GI:11938348
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1140)
AUTHORS NIH-MGC http://mgc.ncbi.nlm.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC) JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI183 row: c column: 04


```

FEATURES
Source
High quality sequence stop: 635.
Location/Qualifiers
1. .1140
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4309563"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 361 a 252 c 277 g 249 t
ORIGIN
Query Match 3.6%; Score 101; DB 12; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 868 CAGGGACGTGTCGAATTGGTCAGTTTGAATGCAACATTTGCCCTACTGGGAATAAC 927
|||||
Db 309 CAGGGACGTGTCGAATTGGTCAGTTTGAATGCAACATTTGCCCTACTGGGAATAAC 368
|||||

QY 928 AGAAGTCGAGGACCTGGGAGCATCTAAAGTGTCAACGTTT 968
|||||
Db 369 AGAAGTCGAGGACCTGGGAGCATCTAAAGTGTCAACGTTT 409
|||||

RESULT 39
AW182255/c
LOCUS
DEFINITION
XJ72r10.x1 Soares_NFL_T_GBC.S1 Homo sapiens cDNA clone
IMAGE:2662795 3' similar to contains Alu repetitive element
; contains element TAR1 repetitive element ; mRNA sequence.
ACCESSION
AW182255
VERSION
AW182255.1 GI:6450715
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 364)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaops-r@mail.nih.gov
This clone is available royalty-free through LML ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 354.
FEATURES
Source
1. .364
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2662795"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCI-CGAP.GC81) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 94 a 86 c 76 g 108 t
ORIGIN
Query Match 3.0%; Score 85; DB 10; Length 364;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1724 AGTTTGAGGAGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATAAC 1783
|||||
Db 85 AGTTTGAGGAGGAGGATGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATAAC 26
|||||

QY 1784 AAGATCCCTGCTCTCAAAAAA 1808
|||||
Db 25 AAGATCCCTGCTCTCAAAAAA 1
|||||

RESULT 40
AA074342/c
LOCUS
DEFINITION
zml5903.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
IMAGE:525748 3' similar to contains Alu repetitive element; mRNA
sequence.
ACCESSION
AA074342
VERSION
AA074342.1 GI:1614210
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 154)
AUTHORS
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissoe, J., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins
, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevas, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
WARNING: There is evidence that suggests that the 384-well parent
plate of this clone contains both human and mouse derived clones.
Thus, the origin of this clone is uncertain. This caution should be
kept in mind should you use this clone.

This clone is available royalty-free through LML ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 126.
FEATURES
Location/Qualifiers
1. .154
/organism="Homo sapiens"
/db_xref="GDB:391717"
/db_xref="taxon:9606"
/clone="IMAGE:525748"
/clone_lib="Stratagene pancreas (#937208)"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Preactic adenocarcinoma cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 28 a 53 c 39 g 34 t

```

```
ORIGIN
Query Match          2.6%; Score 73; DB 9; Length 154;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1983 GCACCTGTAAATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGG 2042
      |||||||
Db 129 GCACCTGTAAATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGG 70
      |||||||

QY 2043 TGGAGGTTGCAGT 2055
      |||||||
Db 69 TGGAGGTTGCAGT 57
      |||||||

RESULT 41
BE988834/c
LOCUS          878 bp      mRNA      linear      EST 29-SEP-2000
DEFINITION    601682124F1 NTH_MGC_9 Homo sapiens cDNA clone IMAGE:3952372 5',
              mRNA sequence.
ACCESSION    BE988834
VERSION      BE988834.1 GI:10365711
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: DCTD/DRP
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
              Plate: LLC821 row: P column: 05
              High quality sequence stop: 783.
              Location/Qualifiers
                1..878
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:3952372"
                  /clone_lib="NTH_MGC_9"
                  /tissue_type="adenocarcinoma cell line"
                  /lab_host="DH10B (phage-resistant)"
                  /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                  EcoRI; cDNA made by oligo-dT priming. Directionally
                  cloned into EcoRI/XhoI sites using the following 5',
                  adaptor: GGCACGAG(G). Size-selected >500bp for average
                  insert size 1.8kb. Library constructed by Ling Hong in
                  the laboratory of Gerald M. Rubin (University of
                  California, Berkeley) using ZAP-cDNA synthesis kit
                  (Stratagene) and Superscript II RT (Life Technologies)."
```

```
BASE COUNT      197 a   257 c   246 g   178 t
ORIGIN

Query Match          2.6%; Score 73; DB 12; Length 878;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GCACCTGTAAATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAG 2041
      |||||||
Db 516 GCACCTGTAAATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAG 457
      |||||||

QY 2042 GTGAGGTTGCAG 2054
      |||||||
Db 456 GTGAGGTTGCAG 444
      |||||||

RESULT 42
AV719004/c
LOCUS          586 bp      mRNA      linear      EST 16-OCT-2000
DEFINITION    AV719004 GLC Homo sapiens cDNA clone GLCEJB05 5', mRNA sequence.
ACCESSION    AV719004
VERSION      AV719004.1 GI:10816156
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 586)
AUTHORS     Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
              , Xu,X., Li,N., Peng,Y., Liu,F., Ou,J., Song,H., Cheng,Z., Zeng,L.,
              Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
              ,Y., Gu,Y., Chen,Z. and Han,Z.
              Homo sapiens cDNA GLC clones
              Unpublished (2000)
              Contact: Zeguang Han
              Chinese National Human Genome Center at Shanghai
              351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
              201203, P. R. China
              Tel: 86-21-50801919(ex.45)
              Fax: 86-21-50801922
              Email: hanzg@hgc.sh.cn
              This clone is available at CHGC in Shanghai.
              Location/Qualifiers
                1..586
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="GLCEJB05"
                  /clone_lib="GLC"
                  /tissue_type="corresponding non cancerous liver tissue"
                  /dev_stage="Adult"
                  /lab_host="SOLR"
                  /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                  XhoI"
              BASE COUNT      139 a   144 c   154 g   149 t
              ORIGIN

Query Match          2.6%; Score 72; DB 10; Length 586;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GCACCTGTAAATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAG 2041
      |||||||
Db 116 GCACCTGTAAATCCCGAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAG 57
      |||||||

QY 2042 GTGAGGTTGCA 2053
      |||||||
Db 56 GTGAGGTTGCA 45
      |||||||

RESULT 43
BE008586/c
LOCUS          252 bp      mRNA      linear      EST 05-JUN-2000
DEFINITION    RC5-BN0158-190400-031-G01 BN0158 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BE008586
VERSION      BE008586.1 GI:8268819
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 252)
AUTHORS     Dias Neto,B., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
              Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
              Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
              Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
              ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
              Simpson,A.J.
              Shotgun sequencing of the human transcriptome with ORF expressed
              sequence tags
```



```

/clone_lib="NCI_CGAP_Prill"
/sex="male"
/tissue_type="normal prostatic epithelial cells"
/lab_host="DH10B"
/notes="Organ: prostate; Vector: pAMP10; mRNA made from
normal prostatic epithelial cells, cDNA made by oligo-dT
priming. Non-directionally cloned. Size-selected on
agarose gel, average insert size 600 bp. Library made by
D. Krizman, NIH."
BASE COUNT      65 a      86 c      70 g      97 t
ORIGIN
Query Match      2.5%; Score 70; DB 9; Length 318;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 2045
Db 137 CCTGTAAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 78
QY 2046 AGGTTGCAGT 2055
Db 77 AGGTTGCAGT 68

RESULT 46
LOCUS AV759459 334 bp mRNA linear EST 19-OCT-2000
DEFINITION AV759459 MDS Homo sapiens cDNA clone MDSAB02 5', mRNA sequence.
ACCESSION AV759459
VERSION AV759459.1 GI:10917307
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 334)
AUTHORS Gu,J., Zhao,M., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,H., Xiao,H.,
Gu,Y., Li,N., Qian,B., Liu,F., Ou,J., Gao,X., Cheng,Z., Xu,Z., Zeng
,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G.,
Yang,Y., Gao,G., Zhang,Q., Chen,S., Han,Z. and Chen,Z.
TITLE Homo sapiens cDNA MDS clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
Location/Qualifiers
1..334
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MDSAB02"
/clone_lib="MDS"
/tissue_type="Bone marrow"
/cell_type="CD34+ hematopoietic stem/progenitor cell"
/lab_host="BM25.8"
/notes="Vector: pTriplEx2; Site_1: sf1IA; Site_2: sf1IB"
BASE COUNT      67 a      103 c      93 g      71 t
ORIGIN
Query Match      2.5%; Score 70; DB 10; Length 334;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 2045
Db 86 CCTGTAAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 145
QY 2046 AGGTTGCAGT 2055

```

```

|||||
Db 146 AGGTTGCAGT 155

RESULT 47
AA211885/c 346 bp mRNA linear EST 13-AUG-1997
LOCUS zr90c03.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:682948 3',
DEFINITION similar to contains Alu repetitive element.; mRNA sequence.
ACCESSION AA211885
VERSION AA211885.1 GI:1810530
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 346)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 260.
FEATURES
Location/Qualifiers
1..346
/organism="Homo sapiens"
/db_xref="GDB:5586268"
/db_xref="taxon:9606"
/clone="IMAGE:682948"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/notes="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI). Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not 1 - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3'.
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not 1 and cloned into the Not 1
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      76 a      90 c      80 g      95 t      5 others
ORIGIN
Query Match      2.5%; Score 70; DB 9; Length 346;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 2045
Db 125 CCTGTAAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 66
QY 2046 AGGTTGCAGT 2055
Db 65 AGGTTGCAGT 56

RESULT 48
LOCUS AV647070 346 bp mRNA linear EST 15-JAN-2002
DEFINITION AV647070 GLC Homo sapiens cDNA clone GLCAUD05 3', mRNA sequence.
ACCESSION AV647070
VERSION AV647070.1 GI:9868084
KEYWORDS EST.
SOURCE human.

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 346)
REFERENCE Xu.X., Huang.J., Xu.Z., Qian.B., Zhu.Z., Yan.Q., Cai.T., Zhang.X.,
AUTHORS Xiao.H., Qu.J., Liu.F., Huang.Q., Cheng.Z., Li.N., Du.J., Hu.W.,
Shen.K., Lu.G., Fu.G., Zhong.M., Xu.S., Gu.W., Huang.W., Zhao.X.,
Hu.G., Gu.J., Chen.Z. and Han.Z.
TITLE Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
MEDLINE 21625106
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@hgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1..346
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCAUD05"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 89 a 77 c 106 g 74 t
ORIGIN

Query Match 2.5%; Score 70; DB 10; Length 346;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCAGCTACTTGGAGCTGAGCAGGAGAGTCTTGAACCCAGGAGGTGG 2045
|||||
Db 205 CCTGTAATCCAGCTACTTGGAGCTGAGCAGGAGAGTCTTGAACCCAGGAGGTGG 264
|||||

QY 2046 AGGTTGCAGT 2055
|||||
Db 265 AGGTTGCAGT 274

RESULT 49
AQ056722 437 bp DNA linear GSS 30-JUL-1998
LOCUS CIT-HSP-233917.TF CIT-HSP Homo sapiens genomic clone 233917, DNA
DEFINITION sequence.
ACCESSION AQ056722
VERSION AQ056722.1 GI:3353328
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 437)
REFERENCE Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
AUTHORS Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-233917.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200

Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

FEATURES
Location/Qualifiers
1..437
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="233917"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 137 a 94 c 118 g 88 t
ORIGIN

Query Match 2.5%; Score 70; DB 17; Length 437;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCAGCTACTTGGAGCTGAGCAGGAGAGTCTTGAACCCAGGAGGTGG 2045
|||||
Db 282 CCTGTAATCCAGCTACTTGGAGCTGAGCAGGAGAGTCTTGAACCCAGGAGGTGG 341
|||||

QY 2046 AGGTTGCAGT 2055
|||||
Db 342 AGGTTGCAGT 351

RESULT 50
N22395/c 448 bp mRNA linear EST 20-DEC-1995
LOCUS yw37f08.s1 Morton Fetal Cochlea Homo sapiens cDNA clone
DEFINITION IMAGE:254439 3' similar to contains Alu repetitive element.; mRNA
sequence.
ACCESSION N22395
VERSION N22395.1 GI:1128529
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 448)
REFERENCE Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
AUTHORS Chissoe,S., Dietrich,N., DuBuque,T., Favellio,A., Gish,W., Hawkins
M., Hultman,M., Kucaba,T., Lacy,M., Le.M., Le.N., Mardis,E., Moore
B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasaki,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 9704478
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 346
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 844 Std Error: 0.00
Seq primer: ml3 -40 forward
High quality sequence stop: 346.
FEATURES
Location/Qualifiers
1..448
/organism="Homo sapiens"
/db_xref="GDB:3891892"

```

/db_xref="taxon:9606"
/clone="IMAGE:254439"
/clone_lib="Morton Petal Cochlea"
/tissue_type="cochlea"
/lab_host="16-22 week fetus"
/notes="Organ: ear; Vector: pBluescript SK-; Site:1: EcoRI;
Site:2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dT. Petal cochlea, normal.
3X of inserts <0.5 Kb, 56% 0.5-1.0 Kb, 7% >1 Kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
-5' adaptor sequence: 5' GAATTCGACACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT      124 a      105 c      94 g      122 t      3 others
ORIGIN

```

```

Query Match      2.5%; Score 70; DB 14; Length 448;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCCGAGCTACTTTGGAGGCTGAGCGAGGAGAAATCGCTTGAACCCAGGAGGTGG 2045
DB 125 CCTGTAATCCCGAGCTACTTTGGAGGCTGAGCGAGGAGAAATCGCTTGAACCCAGGAGGTGG 66

QY 2046 AGGTTGCAGT 2055
DB 65 AGGTTGCAGT 56

```

```

RESULT 51
LOCUS      A0284662
DEFINITION A0284662 544 bp DNA linear GSS 04-MAY-1999
ACCESSION A0284662
VERSION A0284662.1 GI:3910902
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 544)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
TITLE Other_GSSs: RPC111-7809.TJ
JOURNAL
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: 17
Class: BAC ends.
Location/Qualifiers
1..544
/organism="Homo sapiens"
/db_xref="GDB:7525912"
/db_xref="taxon:9606"
/clone="RPCI-11-7809"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/notes="Vector: pBACE3.6; Site:1: EcoRI; Site:2: EcoRI;
RPC11 Human Male BAC Library"

```

```

FEATURES
source

```

```

BASE COUNT      178 a      84 c      133 g      149 t

Query Match      2.5%; Score 70; DB 17; Length 544;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1986 CCTGTAATCCCGAGCTACTTTGGAGGCTGAGCGAGGAGAAATCGCTTGAACCCAGGAGGTGG 2045
DB 439 CCTGTAATCCCGAGCTACTTTGGAGGCTGAGCGAGGAGAAATCGCTTGAACCCAGGAGGTGG 498

QY 2046 AGGTTGCAGT 2055
DB 499 AGGTTGCAGT 508

```

```

RESULT 52
LOCUS      AW973278/c
DEFINITION AW973278 642 bp mRNA linear EST 01-JUN-2000
ACCESSION AW973278
VERSION AW973278.1 GI:8163144
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 642)
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gasparid,R., Gay,C., Holl
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 316
Seq primer: Forward.
Location/Qualifiers
1..642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE_resequences_MAGM"
/notes="Vector: pBluescriptSKm"

```

```

BASE COUNT      167 a      162 c      130 g      183 t

Query Match      2.5%; Score 70; DB 10; Length 642;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1986 CCTGTAATCCCGAGCTACTTTGGAGGCTGAGCGAGGAGAAATCGCTTGAACCCAGGAGGTGG 2045
DB 193 CCTGTAATCCCGAGCTACTTTGGAGGCTGAGCGAGGAGAAATCGCTTGAACCCAGGAGGTGG 134

QY 2046 AGGTTGCAGT 2055
DB 133 AGGTTGCAGT 124

```

```

RESULT 53
LOCUS      A0741937
DEFINITION HS_5569_B2_F07_SP6 RPCI-11 Human Male BAC Library Homo sapiens
ACCESSION A0741937
VERSION A0741937
KEYWORDS GSS.
SOURCE human.

```

```

ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 649)
AUTHORS      Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE        Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL       Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE       99380589
COMMENT      Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University Of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htcsc.washington.edu
Plate: 1145 row: L column: 14
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 649.
FEATURES     source
             Location/Qualifiers
             1..649
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="plate=1145 Col=14 Row=L"
             /clone_lib="RPCI-11 Human Male BAC Library"
             /sex="male"
             /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
             Male blood DNA was isolated from one randomly chosen donor
             and partially digested with a combination of EcoRI and
             EcoRI Methylase. Size selected DNA was cloned into the
             pBACE3.6 vector at EcoRI sites"
BASE COUNT   258 a 88 c 153 g 150 t
ORIGIN
Query Match      2.5%; Score 70; DB 17; Length 649;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCACCTACTTGGGAGGCTGAGCGAGGAGATCGCTTGAACCCAGGAGGTGG 2045
|||||
Db 201 CCTGTAATCCACCTACTTGGGAGGCTGAGCGAGGAGATCGCTTGAACCCAGGAGGTGG 260
|||||
QY 2046 AGGTTGCAGT 2055
|||||
Db 261 AGGTTGCAGT 270

RESULT 54
AG185819
LOCUS         AG185819
DEFINITION   Pan troglodytes DNA, clone: RP43-060A09.TJ, genomic survey
sequence.
ACCESSION    AG185819
VERSION      AG185819.1 GI:16715499
KEYWORDS     GSS.
SOURCE       Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
Male BAC Library
ORGANISM     Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE     1
AUTHORS      Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        BAC end sequences of Library RPCI-43

ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     2 (bases 1 to 705)
AUTHORS      Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL       Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT      Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY      Vector : pBACE3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES     source
             Location/Qualifiers
             1..705
             /organism="Pan troglodytes"
             /db_xref="taxon:9598"
             /clone="RP43-060A09.TJ"
             /sex="male"
             /cell_type="lymphocytes"
             /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT   240 a 138 c 148 g 175 t
ORIGIN
Query Match      2.5%; Score 70; DB 17; Length 705;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCACCTACTTGGGAGGCTGAGCGAGGAGATCGCTTGAACCCAGGAGGTGG 2045
|||||
Db 439 CCTGTAATCCACCTACTTGGGAGGCTGAGCGAGGAGATCGCTTGAACCCAGGAGGTGG 498
|||||
QY 2046 AGGTTGCAGT 2055
|||||
Db 499 AGGTTGCAGT 508

RESULT 55
B14085/c
LOCUS         B14085
DEFINITION   A-837A4.TP CIT978SK Homo sapiens genomic clone A-837A4, DNA
sequence.
ACCESSION    B14085
VERSION      B14085.1 GI:2121834
KEYWORDS     GSS.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 571)
AUTHORS      Adams,M.D., Kelley,J.M., Rounsley,S.R. and Venter,J.C.
TITLE        Use of a BAC End Sequence Database for Sequence-Ready Map Building
JOURNAL       Unpublished (1997)
COMMENT      Other_GSSs: 837A4.TV
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
FEATURES     Location/Qualifiers

```

```
source
1. .571
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="A-837A4"
/clone_lib="CIT978SK"
/sex="Female"
/cell_type="Fibroblast"
/note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
CalTech Human BAC Library A"
BASE COUNT 116 a 167 c 132 g 156 t
ORIGIN

Query Match 2.4%; Score 69; DB 17; Length 571;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1987 CTGTATCCAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGTGG 2046
|||||
Db 222 CTGTAATCCAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGTGG 163
|||||

QY 2047 GGTTCAGT 2055
|||||
Db 162 GGTTCAGT 154

RESULT 56
AQ276581
LOCUS
DEFINITION
AQ276581 300 bp DNA linear GSS 22-NOV-1998
CITBI-EI-2521L5.TR CITBI-EI Homo sapiens genomic clone 2521L5, DNA
sequence.
ACCESSION
AQ276581.1 GI:3902777
VERSION
AQ276581.1
KEYWORDS
GSS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 300)
AUTHORS
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL
Unpublished (1998)
COMMENT
Other_GSSs: CITBI-EI-2521L5.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
Location/Qualifiers
1. .300
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2521L5"
/clone_lib="CITBI-EI"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
CalTech Human BAC Library D"
BASE COUNT 92 a 66 c 84 g 58 t
ORIGIN

Query Match 2.4%; Score 67; DB 17; Length 300;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

source
1. .397
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-436N15"
/clone_lib="RPC111 Human Male BAC Library"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; RPC111 Human Male BAC Library"
BASE COUNT 126 a .81 c 121 g 69 t
ORIGIN

Query Match 2.4%; Score 67; DB 17; Length 397;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGTGG 2045
|||||
Db 205 CCTGTAATCCAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGTGG 264
|||||

QY 2046 AGTTTGC 2052
|||||
Db 265 AGTTTGC 271

RESULT 58
BI023361
LOCUS
DEFINITION
CM4-MT0246-190101-811-d10 MT0246 Homo sapiens cDNA, mRNA sequence.
ACCESSION
BI023361
VERSION
BI023361.1 GI:14429991
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
```


REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 483)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4&t2=CM4-WT0246-190101-811-d10&t3=2001-01-19&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 22
High quality sequence stop: 470.
Location/Qualifiers
1..483
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="WT0246"
/dev_stage="Adult"
/note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 158 a 105 c 128 g 92 t
ORIGIN
Query Match 2.4%; Score 67; DB 13; Length 483;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1986 CCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGATCGTTGAACCCAGGAGGTGG 2045
|||||
Db 387 CCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGATCGTTGAACCCAGGAGGTGG 446
QY 2046 AGGTTGC 2052
|||||
Db 447 AGGTTGC 453
RESULT 59
AQ478179
LOCUS
DEFINITION
725 bp DNA linear GSS 23-APR-1999
RPIC1-11-248K21-TV RPIC1-11 Homo sapiens genomic clone RPIC1-11-248K21
, DNA sequence.
ACCESSION
AQ478179
VERSION
AQ478179.1 GI:4660298
KEYWORDS
GSS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter ,J.C.
Use of BAC End Sequences from Library RPIC1-11 for Sequence-Ready

JOURNAL
COMMENT

Map Building
Unpublished (1997)
Other_GSSs: RPIC1-11-248K21_TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPIC1-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1..725
/organism="Homo sapiens"
/db_xref="GDB:7595108"
/db_xref="taxon:9606"
/clone_lib="RPIC1-11-248K21"
/clone_lib="RPIC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; RPIC11 Human Male BAC Library"
BASE COUNT 227 a 157 c 212 g 129 t
ORIGIN
Query Match 2.4%; Score 67; DB 17; Length 725;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1986 CCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGATCGTTGAACCCAGGAGGTGG 2045
|||||
Db 471 CCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGATCGTTGAACCCAGGAGGTGG 530
QY 2046 AGGTTGC 2052
|||||
Db 531 AGGTTGC 537
RESULT 60
AQ750497/c
LOCUS
DEFINITION
841 bp DNA linear GSS 19-JUL-1999
HS_5576_B1_C06_T7A RPIC1-11 Human Male BAC Library Homo sapiens genomic clone Plate=1152 Col=11 Row=F, DNA sequence.
ACCESSION
AQ750497
VERSION
AQ750497.1 GI:5537655
KEYWORDS
GSS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPIC1-11. For BAC

library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>
 Plate: 1152 row: F column: 11
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 841.

FEATURES
 source
 Location/Qualifiers
 1. .841
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate-1152 Col-11 Row=F"
 /clone_lib="RPC1-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"
 BASE COUNT 200 a 162 c 186 g 291 t 2 others
 ORIGIN

Query Match 2.4%; Score 67; DB 17; Length 841;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAAATCCAGCTACTTGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGGTGG 2045
 |||||||
 Db 540 CCTGTAAATCCAGCTACTTGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGGTGG 481

QY 2046 AGTTGTC 2052
 |||||||
 Db 480 AGTTGTC 474

RESULT 61
 AW882256
 LOCUS 286 bp mRNA linear EST 23-MAY-2000
 DEFINITION RC5-OT0055-080300-021-C12 OT0055 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AW882256
 VERSION AW882256.1 GI:8044266
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 286)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 CONTACT: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=4t2-RC5-OT0055-080300-021-C12&t3=2000-03-08&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 286.

FEATURES
 source
 Location/Qualifiers
 1. .286
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="OT0055"
 /dev_stage="Adult"
 /note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived

FEATURES
 source
 Location/Qualifiers
 1. .286
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="OT0055"
 /dev_stage="Adult"
 /note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived

FEATURES
 source

Location/Qualifiers
 1. .286
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="OT0055"
 /dev_stage="Adult"
 /note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 94 a 62 c 71 g 59 t
 ORIGIN

Query Match 2.3%; Score 66; DB 10; Length 286;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAAATCCAGCTACTTGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGGTGG 2045
 |||||||
 Db 64 CCTGTAAATCCAGCTACTTGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGGTGG 123

QY 2046 AGTTGTC 2051
 |||||||
 Db 124 AGTTGTC 129

RESULT 62
 AW882261
 LOCUS 286 bp mRNA linear EST 23-MAY-2000
 DEFINITION RC5-OT0055-080300-021-D10 OT0055 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AW882261
 VERSION AW882261.1 GI:8044271
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 286)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 CONTACT: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=4t2-RC5-OT0055-080300-021-D10&t3=2000-03-08&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 286.

FEATURES
 source
 Location/Qualifiers
 1. .286
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="OT0055"
 /dev_stage="Adult"
 /note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived

FEATURES
 source
 Location/Qualifiers
 1. .286
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="OT0055"
 /dev_stage="Adult"
 /note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived

FEATURES
 source
 Location/Qualifiers
 1. .286
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="OT0055"
 /dev_stage="Adult"
 /note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived

FEATURES
 source
 Location/Qualifiers
 1. .286
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="OT0055"
 /dev_stage="Adult"
 /note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No. 196 ,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 94 a 62 c 71 g 59 t
ORIGIN

Query Match 2.3%; Score 66; DB 10; Length 286;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAAATCCAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGGTGG 2045
|||||
Db 64 CCTGTAAATCCAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGGTGG 123
|||||

QY 2046 AGGTTG 2051
|||||
Db 124 AGGTTG 129
|||||

RESULT 63

AW882265

LOCUS

RC5-OT0055-080300-021-E12 OT0055 Homo sapiens cDNA, mRNA sequence. EST 23-MAY-2000

ACCESSION

AW882265

VERSION

AW882265.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 286)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC5-OT0055-080

300-021-E12et3-2000-03-08et4-1)

Seq primer: puc 18 forward

High quality sequence stop: 286.

Location/Qualifiers

1. .286

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="OT0055"

/dev_stage="Adult"

/note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2:

SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No. 196

,716 - Ludwig Institute for Cancer Research) profiles

into the pUC 18 vector. Reverse transcription of tissue

mRNA and cDNA amplification were performed under low

stringency conditions."

BASE COUNT 94 a 62 c 71 g 59 t
ORIGIN

Query Match

Best Local Similarity

Matches 66; Conservative

0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAAATCCAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGGTGG 2045

|||||

Db 64 CCTGTAAATCCAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGGTGG 123

|||||

QY 2046 AGGTTG 2051

|||||

Db 124 AGGTTG 129

|||||

RESULT 64

AW882267

LOCUS

RC5-OT0055-080300-021-F12 OT0055 Homo sapiens cDNA, mRNA sequence. EST 23-MAY-2000

ACCESSION

AW882267

VERSION

AW882267.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 286)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC5-OT0055-080

300-021-F12et3-2000-03-08et4-1)

Seq primer: puc 18 forward

High quality sequence stop: 286.

Location/Qualifiers

1. .286

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="OT0055"

/dev_stage="Adult"

/note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2:

SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No. 196

,716 - Ludwig Institute for Cancer Research) profiles

into the pUC 18 vector. Reverse transcription of tissue

mRNA and cDNA amplification were performed under low

stringency conditions."

BASE COUNT 94 a 62 c 71 g 59 t
ORIGIN

Query Match 2.3%; Score 66; DB 10; Length 286;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAAATCCAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGGTGG 2045

|||||

Db 64 CCTGTAAATCCAGCTACTTGGGAGGCTGAGGAGGAGAGATCGCTTGAACCCAGGAGGTGG 123

|||||

```

QY 2046 AGGTTG 2051
Db 124 AGGTTG 129

RESULT 65
AW518030/c
LOCUS
DEFINITION xx65h01.x1 NCI_CGAP_Lym12 Homo sapiens cDNA clone IMAGE:2848561 3'
similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION AW518030
VERSION AW518030.1 GI:7156112
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 388)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Life Technologies catalog #: 11547-015
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
Seq primer: -400P from Gibco
High quality sequence stop: 382.
Location/Qualifiers
1..388
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2848561"
/clone_lib="NCI_CGAP_Lym12"
/tissue_type="lymphoma, follicular mixed small and large
cell"
/lab_host="DH10b"
/notes="Organ: lymph node; Vector: pCMV-Sport6; Site_1:
SalI; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dT. Average insert size 1.25 kb. Life Technologies
catalog #: 11547-015"
96 t 1 others

BASE COUNT 108 a 107 c 76 g 96 t
ORIGIN

Query Match 2.3%; Score 66; DB 10; Length 388;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1990 TAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAAATCGCTTGAACCCAGGAGTGGAGGT 2049
|||||
Db 128 TAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAAATCGCTTGAACCCAGGAGTGGAGGT 69
|||||

QY 2050 TGCAGT 2055
Db 68 TGCAGT 63

RESULT 66
B56190/c
LOCUS
DEFINITION CIT-HSP-2006D8.T8B CIT-HSP Homo sapiens genomic clone 2006D8, DNA
sequence.
ACCESSION B56190
VERSION B56190.1 GI:2610524
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 395)

QY 2046 AGGTTG 2051
Db 124 AGGTTG 129

AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: CIT-HSP-2006D8.T8B
Contact: Mark Adams
Department Of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
Source
1..395
/organism="Homo sapiens"
/db_xref="GDB:7039570"
/db_xref="taxon:9606"
/clone="2006D8"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/notes="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT 91 a 99 c 77 g 128 t
ORIGIN

Query Match 2.3%; Score 66; DB 17; Length 395;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1990 TAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAAATCGCTTGAACCCAGGAGTGGAGGT 2049
|||||
Db 171 TAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAAATCGCTTGAACCCAGGAGTGGAGGT 112
|||||

QY 2050 TGCAGT 2055
Db 111 TGCAGT 106

RESULT 67
AA679794/c
LOCUS
DEFINITION ag92b12.s1 Stratagene hMT neuron (#937233) Homo sapiens cDNA clone
IMAGE:1141919 3' similar to contains Alu repetitive element;; mRNA
sequence.
ACCESSION AA679794
VERSION AA679794.1 GI:2660316
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 402)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theisling,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the

```

IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 339.

FEATURES

source
1..402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1141919"
/clone_lib="Stratagene hNT neuron (#937233)"
/dev_stage="hNT neurons"
/lab_host="SOLR (kanamycin resistant)"
/note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT.
Differentiated, post mitotic hNT neurons. Average insert
size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTATTTTATTTT 3'"

BASE COUNT 78 a 108 c 96 g 120 t

ORIGIN

Query Match 2.3%; Score 66; DB 9; Length 402;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1990 TAATCCAGCTACTTGGGAGGCTGAGCGAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGT 2049
Db 274 TAATCCAGCTACTTGGGAGGCTGAGCGAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGT 215
Qy 2050 TGCAGT 2055
Db 214 TGCAGT 209

RESULT 68

B72122/c 414 bp DNA linear GSS 18-JUN-1998
LOCUS CIT978SK-A-699D2.TPC CIT978SK Homo sapiens genomic clone A-699D2,
DEFINITION DNA sequence.

ACCESSION

B72122

VERSION

B72122.1 GI:2708894

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 414)

AUTHORS

Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden

,K., Berry,K., Granger,D., Suh,E., Wible,C., Kim,U.-J., Shizuya,H.,

Simon,M. and Venter,J.C.

TITLE

Use of a human BAC End Sequence Database for Sequence-Ready Map

JOURNAL

Building

COMMENT

Unpublished (1997)

Other_GSSs: CIT978SK-A-699D2.TVB

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..414

/organism="Homo sapiens"

/db_xref="GDB:1362024"

/db_xref="taxon:9606"

/clone="A-699D2"

/clone_lib="CIT978SK"

/sex="Female"

/cell_type="Fibroblast"
/note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
Caltech Human BAC Library A"

BASE COUNT 97 a 102 c 95 g 120 t

ORIGIN

Query Match 2.3%; Score 66; DB 17; Length 414;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1986 CCTGTAAATCCAGCTACTTGGGAGGCTGAGCGAGGAGAAATCGCTTGAACCCAGGAGGTGG 2045

Db 256 CCTGTAAATCCAGCTACTTGGGAGGCTGAGCGAGGAGAAATCGCTTGAACCCAGGAGGTGG 197

Qy 2046 AGCTTG 2051

Db 196 AGCTTG 191

RESULT 69

AQ111323 430 bp DNA linear GSS 29-AUG-1998
LOCUS CIT-HSP-2371K16.TR CIT-HSP Homo sapiens genomic clone 2371K16, DNA
DEFINITION sequence.

ACCESSION

AQ111323

VERSION

AQ111323.1 GI:3487980

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 430)

AUTHORS

Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,

Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and

Venter,J.C.

TITLE

Use of a random human BAC End Sequence Database for Sequence-Ready

JOURNAL

Map Building

COMMENT

Unpublished (1998)

Other_GSSs: CIT-HSP-2371K16.TF

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: M13 Reverse

Class: BAC ends.

Location/Qualifiers

1..430

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="2371K16"

/clone_lib="CIT-HSP"

/sex="Male"

/cell_type="Sperm"

/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:

HindIII"

BASE COUNT 124 a 89 c 113 g 103 t 1 others

ORIGIN

Query Match 2.3%; Score 66; DB 17; Length 430;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1990 TAATCCAGCTACTTGGGAGGCTGAGCGAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGT 2049

Db 301 TAATCCAGCTACTTGGGAGGCTGAGCGAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGT 360

Qy 2050 TGCAGT 2055

```

Db          361 TGCAGT 366
|||||
RESULT 70
AQ878262/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 434)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 434)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2272 row: F column: 20
Seq primer: T7
Class: BAC ends
High quality sequence stop: 434.
Location/Qualifiers
1..434
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Ov23"
/tissue_type="tumor, 5 pooled (see description)"
/lab_host="DH10B"
/notes="Organ: ovary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.35 kb. Tumor types include: mixed
Mullerian tumor, papillary serous, clear cell, spindle
cell. All are primary tumors, metastasis positive. Life
Technologies catalog #: 11534-013"
BASE COUNT 91 a 109 c 92 g 140 t 2 others
ORIGIN
Query Match 2.3%; Score 66; DB 17; Length 434;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1986 CCTGATATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 2045
|||||
Db 238 CCTGATATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGG 179
Qy 2046 AGGTTG 2051
|||||
Db 178 AGGTTG 173
RESULT 71
AI653493/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 526)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 526)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Other_ESTs: ba84g06.y1
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1185 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 417.
Location/Qualifiers
1..451
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Ov23"
/tissue_type="tumor, 5 pooled (see description)"
/lab_host="DH10B"
/notes="Organ: ovary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.35 kb. Tumor types include: mixed
Mullerian tumor, papillary serous, clear cell, spindle
cell. All are primary tumors, metastasis positive. Life
Technologies catalog #: 11534-013"
BASE COUNT 133 a 121 c 83 g 114 t
ORIGIN
Query Match 2.3%; Score 66; DB 9; Length 451;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1990 TAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGT 2049
|||||
Db 127 TAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGT 68
Qy 2050 TGCAGT 2055
|||||
Db 67 TGCAGT 62
RESULT 72
BE301068/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 526)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Other_ESTs: ba84g06.y1
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1185 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 417.
Location/Qualifiers
1..451
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Ov23"
/tissue_type="tumor, 5 pooled (see description)"
/lab_host="DH10B"
/notes="Organ: ovary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.35 kb. Tumor types include: mixed
Mullerian tumor, papillary serous, clear cell, spindle
cell. All are primary tumors, metastasis positive. Life
Technologies catalog #: 11534-013"
BASE COUNT 133 a 121 c 83 g 114 t
ORIGIN
Query Match 2.3%; Score 66; DB 9; Length 451;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1990 TAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGT 2049
|||||
Db 127 TAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGT 68
Qy 2050 TGCAGT 2055
|||||
Db 67 TGCAGT 62
RESULT 72
BE301068/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 526)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Other_ESTs: ba84g06.y1
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1185 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 417.
Location/Qualifiers
1..451
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Ov23"
/tissue_type="tumor, 5 pooled (see description)"
/lab_host="DH10B"
/notes="Organ: ovary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.35 kb. Tumor types include: mixed
Mullerian tumor, papillary serous, clear cell, spindle
cell. All are primary tumors, metastasis positive. Life
Technologies catalog #: 11534-013"
BASE COUNT 133 a 121 c 83 g 114 t
ORIGIN
Query Match 2.3%; Score 66; DB 9; Length 451;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1990 TAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGT 2049
|||||
Db 127 TAATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGT 68
Qy 2050 TGCAGT 2055
|||||
Db 67 TGCAGT 62

```

found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
Seq primer: -40UP from Gibco
High quality sequence stop: 441.

FEATURES

source

1..526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2907130"
/clone_lib="NIH_MGC_21"
/tissue="type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT
ORIGIN

164 a 134 c .96 g 132 t

Query Match 2.3%; Score 66; DB 10; Length 526;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1990 TAATCCAGCTACTTGGGAGGCTGAGCAGGAGAGATCGCTTGAACCCAGGAGGTGGAGCT 2049
|||||

Db

129 TAATCCAGCTACTTGGGAGGCTGAGCAGGAGATCGCTTGAACCCAGGAGGTGGAGCT 70
|||||

QY

2050 TGCAGT 2055

Db

69 TGCAGT 64

RESULT 73

BE156416

LOCUS

QV0-HT0368-030100-080-c12 HT0368 Homo sapiens cDNA, mRNA linear EST 21-JUN-2000

DEFINITION

BE156416

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=QV0-HT0368-030

100-080-c12&t3=2000-01-03&t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 395.

FEATURES

source

Location/Qualifiers

source

1..395
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0368"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT
ORIGIN

117 a 79 c 115 g 84 t

Query Match 2.3%; Score 65; DB 10; Length 395;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1991 AATCCAGCTACTTGGGAGGCTGAGCAGGAGAGATCGCTTGAACCCAGGAGGTGGAGTT 2050
|||||

Db

236 AATCCAGCTACTTGGGAGGCTGAGCAGGAGAGATCGCTTGAACCCAGGAGGTGGAGTT 295
|||||

QY

2051 GCAGT 2055

Db

296 GCAGT 300

RESULT 74

BE156526

LOCUS

QV0-HT0368-310100-091-c07 HT0368 Homo sapiens cDNA, mRNA linear EST 21-JUN-2000

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

CONTACT

LABORATORY

INSTITUTE

FOR

RESEARCH

LOCATION

QUALIFIERS

FEATURES

source

Location/Qualifiers

1..400

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="HT0368"

/dev_stage="Adult"

/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 118 a 80 c 115 g 87 t

ORIGIN

Query Match 2.3%; Score 65; DB 10; Length 400;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1991 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGGAGGTGAGGTT 2050

Db 241 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGGAGGTGAGGTT 300

QY 2051 GCAGT 2055

Db 301 GCAGT 305

RESULT 75

BE940056

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

Source

1..443

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="UT0034"

/dev_stage="Adult"

/note="Organ: uterus; tumor: Vector: puc18; Site_1: SmaI;

Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 132 a 90 c 134 g 87 t

ORIGIN

Query Match 2.3%; Score 65; DB 10; Length 400;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1991 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGGAGGTGAGGTT 2050

Db 241 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGGAGGTGAGGTT 300

QY 2051 GCAGT 2055

Db 301 GCAGT 305

RESULT 75

BE940056

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

Source

1..443

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="UT0034"

/dev_stage="Adult"

/note="Organ: uterus; tumor: Vector: puc18; Site_1: SmaI;

Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 132 a 90 c 134 g 87 t

ORIGIN

Query Match 2.3%; Score 65; DB 10; Length 400;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1991 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGGAGGTGAGGTT 2050

Db 241 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGGAGGTGAGGTT 300

QY 2051 GCAGT 2055

Db 301 GCAGT 305

RESULT 75

BE940056

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

Source

1..443

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="UT0034"

/dev_stage="Adult"

/note="Organ: uterus; tumor: Vector: puc18; Site_1: SmaI;

Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 132 a 90 c 134 g 87 t

ORIGIN

Query Match 2.3%; Score 65; DB 10; Length 400;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1991 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGGAGGTGAGGTT 2050

Db 241 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGGAGGTGAGGTT 300

QY 2051 GCAGT 2055

Db 301 GCAGT 305

RESULT 75

BE940056

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

Source

1..443

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="UT0034"

/dev_stage="Adult"

/note="Organ: uterus; tumor: Vector: puc18; Site_1: SmaI;

Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 132 a 90 c 134 g 87 t

ORIGIN

Query Match 2.3%; Score 65; DB 10; Length 400;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1991 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGGAGGTGAGGTT 2050

Db 241 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGGAGGTGAGGTT 300

QY 2051 GCAGT 2055

Db 301 GCAGT 305

RESULT 75

BE940056

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

Source

1..443

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="UT0034"

/dev_stage="Adult"

/note="Organ: uterus; tumor: Vector: puc18; Site_1: SmaI;

Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 132 a 90 c 134 g 87 t

ORIGIN

Query Match 2.3%; Score 65; DB 10; Length 400;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1991 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGGAGGTGAGGTT 2050

Db 241 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGGAGGTGAGGTT 300

QY 2051 GCAGT 2055

Db 301 GCAGT 305

RESULT 75

BE940056

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

Source

1..443

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="UT0034"

/dev_stage="Adult"

/note="Organ: uterus; tumor: Vector: puc18; Site_1: SmaI;

Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 132 a 90 c 134 g 87 t

ORIGIN

Query Match 2.3%; Score 65; DB 10; Length 400;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1991 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGGAGGTGAGGTT 2050

Db 241 AATCCAGCTACTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGGAGGTGAGGTT 300

QY 2051 GCAGT 2055

Db 301 GCAGT 305

RESULT 75

BE940056

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

Source

BG059756/c
LOCUS BG059756 468 bp mRNA linear EST 25-JAN-2001
DEFINITION naf53c05.x1 NCI_CGAP_Brn65 Homo sapiens cDNA clone IMAGE:4147593 3'
similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION BG059756
VERSION BG059756.1 GI:12527556
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 468)
AUTHORS NCI/NINDS-CoAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 416.
FEATURES
Location/Qualifiers
1..468
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4147593"
/clone_lib="NCI_CGAP_Brn65"
/tissue_type="glioblastoma without EGFR amplification"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 1.77 kb. Constructed by Life Technologies."
BASE COUNT 92 a 135 c 104 g 136 t 1 others
ORIGIN
1
1111
Query Match 2.3%; Score 64; DB 12; Length 468;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1983 GCACCTGTATCCACGCTACTTGGGAGGCTGAGCAGGAGATCGCTTGAACCCAGGAGG 2042
|||||
Db 127 GCACCTGTATCCACGCTACTTGGGAGGCTGAGCAGGAGATCGCTTGAACCCAGGAGG 68
QY 2043 TCGA 2046
|||||
Db 67 TCGA 64
RESULT 78
AQ344499
LOCUS AQ344499
DEFINITION RPC111-13402.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-13402,
DNA sequence.
ACCESSION AQ344499
VERSION AQ344499.1 GI:4169395
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 470)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building

JOURNAL Unpublished (1997)
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
FEATURES
Location/Qualifiers
1..470
/organism="Homo sapiens"
/db_xref="GDB:7551409"
/db_xref="taxon:9606"
/clone="RPCI-11-13402"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPC111 Human Male BAC Library"
BASE COUNT 121 a 100 c 134 g 115 t
ORIGIN
1
1111
Query Match 2.3%; Score 64; DB 17; Length 470;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1979 GCAGGCACCTGTAAATCCAGCTACTTGGAGGCTGAGGAGGAGATCGCTTGAACCCAG 2038
|||||
Db 193 GCAGGCACCTGTAAATCCAGCTACTTGGAGGCTGAGGAGGAGATCGCTTGAACCCAG 252
QY 2039 GAGG 2042
|||||
Db 253 GAGG 256
RESULT 79
N49425/c
LOCUS N49425 479 bp mRNA linear EST 14-FEB-1996
DEFINITION yv21c11.r1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone
IMAGE:243380 5' similar to contains Alu repetitive element;; mRNA
sequence.
ACCESSION N49425
VERSION N49425.1 GI:1190591
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 479)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisakis,E., Waterston,R., Williamson,A., Wohldmann,P. and Willson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: T7
High quality sequence stop: 357.

	/db_xref=""taxon:9606" /clone=""RPCI-11-259A5" /clone_lib=""RPCI-11" /sex=""Male" /cell_type=""Lymphocytes" /note=""Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPChII Human Male BAC Library"	BASE COUNT ORIGIN	145 a 175 c 150 g 187 t
	Query Match 2.3%; Score 64; DB 17; Length 657;	Best Local Similarity 100.0%; Pred. No. 0;	Mismatches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1979 GCAGGCACCTGTAAATCCCGACTTGGGAGCCTTGAGCGAGAGAATCGTTGAACCAG 2038 DB	179 GCAGGCACCTGTAAATCCCGACTTGGGAGCCTTGAGCGAGAGAATCGTTGAACCAG 238 	QY 2039 GAGG 2042
Db			239 GAGG 242
RESULT 81			
LOCUS	AG041398/c		
DEFINITION	Pan troglodytes DNA, clone: PTB-019B13.R, genomic survey sequence.		
ACCESSION	AG041398		
VERSION	AG041398.1 GI:16570123		
KEYWORDS	GSS.		
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male		
ORGANISM	BAC Library clone:PTB-019B13.R. Pan troglodytes Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.		
REFERENCE	1		
AUTHORS	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.		
TITLE	BAC end sequences of Library PTB		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 665)		
AUTHORS	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (Riken), Genomic Sciences Center (GSC); 1-7-2 Suhiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimbges@gsc.riken.go.jp, URL:http://hgps.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170) Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.		
COMMENT	PRIMERS Sequencing: MJ3Rev LIBRARY Vector : pKS145 R.Site 1 : SacI R.Site 2 : SacI. Location/Qualifiers 1..665 /organism=""Pan troglodytes" /db_xref=""taxon:9598" /clone=""PTB-019B13.R" /sex=""male" /cell_type=""lymphoblast" /clone_lib=""PTB Chimpanzee Male BAC Library" 		
FEATURES	source		
BASE COUNT	140 a 174 c 157 g 193 t		
ORIGIN			
Query Match 2.3%; Score 64; DB 17; Length 665;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1979 GCAGGCACCTGTAAATCCCGACTTGGGAGCCTTGAGCGAGAGAATCGTTGAACCAG 2038 DB	311 GCAGGCACCTGTAAATCCCGACTTGGGAGCCTTGAGCGAGAGAATCGTTGAACCAG 252 	QY 2039 GAGG 2042
Db			251 GAGG 248
RESULT 80			
LOCUS	AQ480012		
DEFINITION	RPCI-11-259A5.TV RPCI-11 Homo sapiens genomic clone RPCI-11-259A5, DNA sequence.		
ACCESSION	AQ480012		
VERSION	AQ480012.1 GI:4662131		
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 657)		
AUTHORS	Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter J.C.		
TITLE	Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building		
JOURNAL	Unpublished (1997)		
COMMENT	Other.GSSs: RPCI-11-259A5.TJ Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208 Email: hbe@tigr.org Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html . Seq primer: T7 Class: BAC ends.		
FEATURES	location/qualifiers		
source	1..657 /organism=""Homo sapiens" /db_xref=""GDB:7599076"		

```

Qy 1979 GCAGGCACCTGTATCCAGCTACTTGGGAGGCTGAGGCAGGAGAAATCGTTGAACCCAG 2038
      |||||||
Db 266 GCAGGCACCTGTATCCAGCTACTTGGGAGGCTGAGGCAGGAGAAATCGTTGAACCCAG 207
      |||||||
Qy 2039 GAGG 2042
      |||||
Db 206 GAGG 203

RESULT 82
AZ516708
LOCUS
DEFINITION
  A2516708 263 bp DNA linear GSS 16-OCT-2000
  RPCI-11-195F6.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-195F6,
  DNA sequence.
ACCESSION
  A2516708
VERSION
  A2516708.1 GI:10825382
KEYWORDS
  GSS.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 263)
  Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
  ,J.C.
  BAC end sequences of library RPCI-11
  Unpublished (1997)
  Other_GSSs: RPCI-11-195F6.TV
  Contact: Shaying Zhao
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: szhao@tigr.org
  Clones are derived from the human BAC library RPCI-11. For BAC
  library availability, please contact Pieter de Jong
  (pieter@dejong.med.buffalo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
  Research Genet cs (info@resgen.com). BAC end search page:
  http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
  This BAC end was generated during the R&D process and may have
  higher chance of clone tracking errors.
  Seq primer: SP6
  Class: BAC ends.
FEATURES
  source
    1..263
    /organism="Homo sapiens"
    /db_xref="GDB:7574521"
    /db_xref="taxon:9606"
    /clone="RPCI-11-195F6"
    /clone_lib="RPCI-11"
    /sex="Male"
    /cell_type="Lymphocytes"
    /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
    RPCI11 Human Male BAC Library"
  BASE COUNT
    91 a 60 c 62 g 49 t 1 others
  ORIGIN
    Query Match 2.2%; Score 63; DB 17; Length 263;
    Best Local Similarity 100.0%; Pred. No. 0;
    Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1974 TGGTAGCAGGCACCTGTATCCAGCTACTTGGGAGGCTGAGGCAGGAGAAATCGCTTGAA 2033
      |||||||
Db 36 TGGTAGCAGGCACCTGTATCCAGCTACTTGGGAGGCTGAGGCAGGAGAAATCGCTTGAA 95
      |||||
Qy 2034 CCC 2036
      |||||
Db 96 CCC 98

RESULT 83
BF933309/c

```

```

LOCUS
DEFINITION
  BF933309 380 bp mRNA linear EST 22-JAN-2001
  IL5-NT0273-201200-369-e03 NT0273 Homo sapiens cDNA, mRNA sequence.
ACCESSION
  BF933309
VERSION
  BF933309.1 GI:12350633
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 380)
  Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
  Negai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
  Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
  Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
  ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
  Simpson,A.J.
  Shotgun sequencing of the human transcriptome with ORF expressed
  sequence tags
  Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  20202663
  Contact: Simpson A.J.G.
  Laboratory of Cancer Genetics
  Ludwig Institute for Cancer Research
  Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
  Brazil
  Tel: +55-11-2704922
  Fax: +55-11-2707001
  Email: asimpson@ludwig.org.br
  This sequence was derived from the FAPESP/LICR Human Cancer Genome
  Project. This entry can be seen in the following URL
  (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&t2=IL5-NT0273-
  201200-369-e03&t3=2000-12-20&t4=1)
  Seq primer: puc 18 forward
  High quality sequence stop: 342.
FEATURES
  source
    1..380
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="NT0273"
    /dev_stage="Adult"
    /note="Organ: nervous tumor; Vector: puc18; Site_1: SmaI;
    Site_2: SmaI; A mini-library was made by cloning products
    derived from ORESTES PCR (U.S. Letters Patent application
    No. 196,716 - Ludwig Institute for Cancer Research)
    profiles into the pUC 18 vector. Reverse transcription of
    tissue mRNA and cDNA amplification were performed under
    low stringency conditions."
  BASE COUNT
    77 a 91 c 92 g 119 t 1 others
  ORIGIN
    Query Match 2.2%; Score 63; DB 12; Length 380;
    Best Local Similarity 100.0%; Pred. No. 0;
    Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1993 TCCCAGCTACTTGGGAGGCTGAGGCAGGAGAAATCGTTGAACCCAGGAGGTGGC 2052
      |||||||
Db 193 TCCCAGCTACTTGGGAGGCTGAGGCAGGAGAAATCGTTGAACCCAGGAGGTGGC 134
      |||||||
Qy 2053 AGT 2055
      |||||
Db 133 AGT 131

RESULT 84
AI673749/c
LOCUS
DEFINITION
  AI673749 424 bp mRNA linear EST 15-DEC-1999
  tw79ell.x1 NCI_CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2265932.3,
  similar to contains Alu repetitive element; contains element THR
  repetitive element ;, mRNA sequence.
ACCESSION
  AI673749
VERSION
  AI673749.1 GI:4853480
KEYWORDS
  EST.
SOURCE
  human.

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 424)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 2630 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 411.
Location/Qualifiers
1..424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clones="IMAGE:2265932"
/clone_lib="NCI-CGAP_Ut3"
/tissue_types="poorly-differentiated endometrial adenocarcinoma 2 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.45 kb. Life Technologies catalog #: 11541-018"

BASE COUNT 93 a 126 c 96 g 109 t
ORIGIN

Query Match 2.2%; Score 63; DB 9; Length 424;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGGCACCTGTAATCCAGCTACTTGGGAGGCTGAGGAGGAGATCGCTTGAACCCAG 2038
|||||
Db 140 GCAGGCACCTGTAATCCAGCTACTTGGGAGGCTGAGGAGGAGATCGCTTGAACCCAG 81
|||||

QY 2039 GAG 2041
|||
Db 80 GAG 78

RESULT 85
AQ238522
LOCUS AQ238522 445 bp DNA linear GSS 21-APR-1999
DEFINITION RPC111-70N2.TJ RPC1-11 Homo sapiens genomic clone RPC1-11-70N2, DNA sequence.
ACCESSION AQ238522
VERSION AQ238522.1 GI:3670813
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 445)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Barry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1..445
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Barry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: mdadams@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://inforesgen.com>). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..445
/organism="Homo sapiens"
/db_xref="GDB:7526809"
/db_xref="taxon:9606"
/clone="RPC1-11-70M2"
/clone_lib="RPC1-11"
/cell_type="Lymphocytes"
/sex="Male"
/notes="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPC111 Human Male BAC Library"

BASE COUNT 167 a 94 c 94 g 90 t
ORIGIN

Query Match 2.2%; Score 63; DB 17; Length 445;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1974 TGTTAGCAGGCACCTGTAATCCAGCTACTTGGGAGGCTGAGGAGGAGATCGCTTGA 2033
|||||
Db 131 TGTTAGCAGGCACCTGTAATCCAGCTACTTGGGAGGCTGAGGAGGAGATCGCTTGA 190
|||||

QY 2034 CCC 2036
|||
Db 191 CCC 193

RESULT 86
AQ268452
LOCUS AQ268452 454 bp DNA linear GSS 27-APR-1999
DEFINITION RPC111-69015.TJ RPC1-11 Homo sapiens genomic clone KPCI-11-69015, DNA sequence.
ACCESSION AQ268452
VERSION AQ268452.1 GI:3796056
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 454)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Barry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: RPC111-69015.TK
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://inforesgen.com>). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..454
/organism="Homo sapiens"
/db_xref="GDB:7526462"

```

/db_xref="taxon:9606"
/clone="RPCI-11-69015"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT      169 a   98 c  102 g   85 t
ORIGIN

Query Match      2.2%; Score 63; DB 17; Length 454;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1974 TGGTAGCAGGCACCTGTAAATCCAGCTACTTTGGGAGGCTGAGCAGGAGAAATCGCTTGAA 2033
|||||
Db 148 TGGTAGCAGGCACCTGTAAATCCAGCTACTTTGGGAGGCTGAGCAGGAGAAATCGCTTGAA 207
|||||

Qy 2034 CCC 2036
|||
Db 208 CCC 210

RESULT 87
A0017563/c
LOCUS      A0017563      490 bp      DNA      linear      GSS 09-JUN-1998
DEFINITION CIT-HSP-2306J3.TF CIT-HSP Homo sapiens genomic clone 2306J3, DNA
sequence.
ACCESSION  A0017563
VERSION    A0017563.1 GI:3196299
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 490)
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL   Unpublished (1998)
COMMENT   Other_GSSs: CIT-HSP-2306J3.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..490
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone="2306J3"
     /clone_lib="CIT-HSP"
     /sex="Male"
     /cell_type="Sperm"
     /note="vector: pBAC3.6; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT      94 a   135 c   110 g   151 t
ORIGIN

Query Match      2.2%; Score 63; DB 17; Length 490;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1993 TCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGTGC 2052
|||||

```

```

Db 157 TCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGTGC 98
|||||
Qy 2053 AGT 2055
|||
Db 97 AGT 95

RESULT 88
A0624960
LOCUS      A0624960      502 bp      DNA      linear      GSS 16-JUN-1999
DEFINITION CITBI-EI-2649E20.TF CITBI-EI Homo sapiens genomic clone 2649E20,
DNA sequence.
ACCESSION  A0624960
VERSION    A0624960.1 GI:5087352
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 502)
AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
Map Building
JOURNAL   Unpublished (1997)
COMMENT   Other_GSSs: CITBI-EI-2649E20.TR
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeet@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..502
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone="2649E20"
     /clone_lib="CITBI-EI"
     /sex="male"
     /cell_type="sperm"
     /note="vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"
BASE COUNT      150 a   100 c   123 g   129 t
ORIGIN

Query Match      2.2%; Score 63; DB 17; Length 502;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1993 TCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGTGC 2052
|||||
Db 319 TCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGTGC 378
|||||
Qy 2053 AGT 2055
|||
Db 379 AGT 381

RESULT 89
AG127767
LOCUS      AG127767      639 bp      DNA      linear      GSS 04-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-138116.F, genomic survey sequence.
ACCESSION  AG127767
VERSION    AG127767.1 GI:16656932
KEYWORDS   GSS.
SOURCE     Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male

```



```

BASE COUNT      152 a   167 c   141 g   255 t   1 others
ORIGIN

Query Match      2.2%; Score 63; DB 9; Length 716;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1993 TCCAGCTACTTGGGAGGCTGAGGCAGGAGATCGCTTGAACCCAGGAGGTTC 2052
    |||||
Db 135 TCCAGCTACTTGGGAGGCTGAGGCAGGAGATCGCTTGAACCCAGGAGGTTC 76
    |||||

QY 2053 AGT 2055
    |||
Db 75 AGT 73

RESULT 92
BQ707295
LOCUS      734 bp      mRNA      linear      EST 16-JUL-2002
DEFINITION AGENCOURT_8475232 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301509
5', mRNA sequence.
ACCESSION  BQ707295
VERSION     BQ707295
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 734)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Dr. Mark Watson
            cDNA Library Preparation: Rubin Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CM2517 row: h column: 22
            High quality sequence stop: 606.
            Location/Qualifiers
            1. .734
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="NIH_MGC_113"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGCACGAG(G). Library constructed by Ling Hong in the
            laboratory of Gerald M. Rubin (University of California,
            Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
            Superscript II RT (Life Technologies). Note: this is a
            NIH_MGC Library."
            BASE COUNT      227 a   164 c   196 g   146 t   1 others
            ORIGIN

Query Match      2.2%; Score 63; DB 14; Length 734;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGGCACCTGTAATCCAGCTACTTGGGAGGCTGAGGAGGAGATCGCTTGAACCCAG 2038
    |||||
Db 586 GCAGGCACCTGTAATCCAGCTACTTGGGAGGCTGAGGAGGAGATCGCTTGAACCCAG 645
    |||||

QY 2039 GAG 2041
    |||
Db 646 GAG 648

```

```

RESULT 93
BQ755005
LOCUS      1034 bp      mRNA      linear      EST 15-MAY-2001
DEFINITION 602711311F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4851819 5',
mRNA sequence.
ACCESSION  BQ755005
VERSION     BQ755005
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 1034)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CM1695 row: e column: 04
            High quality sequence stop: 622.
            Location/Qualifiers
            1. .1034
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="NIH_MGC_48"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
            Site_2: EcoRI; cDNA made by oligo-dT priming.
            Directionally cloned into EcoRI/XhoI sites using the
            following 5' adaptor: GGCACGAG(G). Size-selected >500bp
            for average insert size 1.8Kb. Library constructed by Ling
            Hong in the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies).
            Note: this is a NIH_MGC Library."
            BASE COUNT      305 a   211 c   233 g   283 t   2 others
            ORIGIN

Query Match      2.2%; Score 62; DB 12; Length 1034;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 919 GGGATAACAGAACTGCAGGACCTGGGACCATCTAAAGTCATCAGGTTTCTATGACT 978
    |||||
Db 457 GGGATAACAGAACTGCAGGACCTGGGACCATCTAAAGTCATCAGGTTTCTATGACT 516
    |||||

QY 979 TT 980
    ||
Db 517 TT 518

```

```

RESULT 94
AA993041/c
LOCUS      308 bp      mRNA      linear      EST 27-AUG-1998
DEFINITION ct92f06.s1 Soares_total_fetus_Nb2HF8.9w Homo sapiens cDNA clone
IMAGE:1624259 3' similar to contains Alu repetitive element; , mRNA
sequence.
ACCESSION  AA993041
VERSION     AA993041
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 308)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 731 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 299.

FEATURES
source
1..308
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1624259"
/clone_lib="Soares_total_fetus_Nb2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTACAAATCTGAAGTGGAGCGCGCTTAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 74 a 72 c 82 g 80 t
ORIGIN

Query Match 2.2%; Score 61; DB 9; Length 308;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTGTAAATCCCGAGCTACTTGGGAGGCTGAGGAGAGAAATCGCTTGAACCCAGGAG 2041
|||||
Db 137 GGCACCTGTAAATCCCGAGCTACTTGGGAGGCTGAGGAGAGAAATCGCTTGAACCCAGGAG 78
|||||

QY 2042 G 2042
|
Db 77 G 77

RESULT 95
BM988047/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BM988047
UI-H-DF0-atq-k-14-0-UI.s2 NCI_CGAP_DFO Homo sapiens cDNA clone
IMAGE: 3108204 3', mRNA sequence.
BM988047
BM988047.1 GI:19707436
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 396)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: NCI-CGAP clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
The following repetitive elements were found in this cDNA

sequence: 11-296, >ALU (matched complement) 289-351.
>MER5A#DNA/MER1_type
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source
1..396
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE: 3108204"
/clone_lib="NCI_CGAP_DFO"
/tissue_type="Subchondral Bone"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="organ: Bone; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI_CGAP_DFO is a cDNA library containing the following
tissue(s): Subchondral Bone. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (GT)18 tail. The
sequence tag for this library is GTTAAGCGTC.
TAG_L1B=UI-H-DF0
TAG_TISSUE=subchondral bone
TAG_SEQ=GTTAAGCGTC"

BASE COUNT 74 a 107 c 83 g 132 t
ORIGIN

Query Match 2.2%; Score 61; DB 14; Length 396;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTGTAAATCCCGAGCTACTTGGGAGGCTGAGGAGAGAAATCGCTTGAACCCAGGAG 2041
|||||
Db 149 GGCACCTGTAAATCCCGAGCTACTTGGGAGGCTGAGGAGAGAAATCGCTTGAACCCAGGAG 90
|||||

QY 2042 G 2042
|
Db 89 G 89

RESULT 96
AA838140/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AA838140
cellb04.s1 NCI_CGAP_Ov2 Homo sapiens cDNA clone IMAGE:1385551
similar to contains Alu repetitive element; mRNA sequence.
AA838140
AA838140.1 GI:2913797
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 400)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
[www-bio.llnl.gov/bbrp/image/image.html](http://image.llnl.gov/bbrp/image/image.html)
Insert Length: 1249 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 353.

FEATURES

SOURCE

1. .400
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1385551"
/clone_lib="NCI_CGAP_Ov2"
/sex="female"
/tissue_type="ovary"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from invasive ovarian tumor, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."

BASE COUNT 74 a 106 c 102 g 118 t

ORIGIN

Query Match 2.2%; Score 61; DB 9; Length 400;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTGTAAATCCAGCTACTTGGAGGCTGAGCAGAGAAATCGCTTGAACCCAGGAG 2041
|||||
Db 136 GGCACCTGTAAATCCAGCTACTTGGAGGCTGAGCAGAGAAATCGCTTGAACCCAGGAG 77

QY 2042 G 2042

Db 76 G 76

RESULT 97

AQ596495/c

LOCUS

HS_5191_AL_D12_SP6E RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=767 Col=23 Row=G, DNA sequence.

ACCESSION

AQ596495

VERSION

AQ596495.1

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 404)

AUTHORS

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE

Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

MEDLINE

99380589

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu

Plate: 767 row: G column: 23

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 404.

FEATURES

SOURCE

1. .404
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone="Plate=767 Col=23 Row=G"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"

/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"

BASE COUNT 104 a 86 c 77 g 137 t

ORIGIN

Query Match 2.2%; Score 61; DB 17; Length 404;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1995 CCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGTTGCGAG 2054
|||||
Db 359 CCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGTTGCGAG 300

QY 2055 T 2055

Db 299 T 299

RESULT 98

AQ414005/c

LOCUS

AQ414005 430 bp DNA linear GSS 23-MAR-1999
RPCI-11-194F22.TV RPCI-11 Homo sapiens genomic clone RPCI-11-194F22
, DNA sequence.

ACCESSION

AQ414005

VERSION

AQ414005.1

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 430)

AUTHORS

Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.

TITLE

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building

JOURNAL

Unpublished (1997)

COMMENT

Other_GSSs: RPCI-11-194F22.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html.
Seq primer: T7

FEATURES

source

1. .430
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:7574253"
/db_xref="taxon:9606"
/clone="RPCI-11-194F22"
/clone_lib="RPCI-11"
/sex="Male"

/cell_type="Lymphocytes"

/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;

RPCI11 Human Male BAC Library"

BASE COUNT 111 a 92 c 86 g 141 t

ORIGIN

Query Match 2.2%; Score 61; DB 17; Length 430;

Search completed: July 17, 2003, 23:12:39
Job time : 3603 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2003, 21:32:20 ; Search time 7088 seconds
(without alignments)
11570.500 Million cell updates/sec

Title: US-09-966-880A-7
Perfect score: 2818
Sequence: 1 agagaccatcattatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database :

GenEmbl.*				
1:	gb_ba.*			
2:	gb_htg.*			
3:	gb_in.*			
4:	gb_om.*			
5:	gb_ov.*			
6:	gb_pat.*			
7:	gb_ph.*			
8:	gb_pl.*			
9:	gb_pr.*			
10:	gb_ro.*			
11:	gb_sts.*			
12:	gb_sy.*			
13:	gb_un.*			
14:	gb_vi.*			
15:	em_ba.*			
16:	em_fun.*			
17:	em_hum.*			
18:	em_in.*			
19:	em_mu.*			
20:	em_om.*			
21:	em_or.*			
22:	em_ov.*			
23:	em_pat.*			
24:	em_ph.*			
25:	em_pi.*			
26:	em_ro.*			
27:	em_sts.*			
28:	em_un.*			
29:	em_vi.*			
30:	em_htg_hum.*			
31:	em_htg_inv.*			
32:	em_htg_other.*			
33:	em_htg_mus.*			
34:	em_htg_pln.*			
35:	em_htg_rtd.*			
36:	em_htg_mam.*			
37:	em_htg_vrt.*			
38:	em_sy.*			
39:	em_htgo_hum.*			
40:	em_htgo_mus.*			
41:	em_htgo_other.*			

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2791	99.0	2791	9	AB040431 Homo sapi
2	2174	77.1	11204	9	AB040430 Homo sapi
3	2174	77.1	71132	9	AC092184 Homo sapi
4	1815	64.4	1837	9	BC006296 Homo sapi
5	587	20.8	596	12	AF529829 Mus muscu
6	584	20.7	596	9	AF529823 Homo sapi
7	572	20.3	597	12	AF529846 Crictetulu
8	569	20.2	597	12	AF529842 Crictetulu
9	548	19.4	597	12	AF529847 Crictetulu
10	546	19.4	597	12	AF529828 Mus muscu
11	545	19.3	596	9	AF529815 Homo sapi
12	545	19.3	596	9	AF529816 Homo sapi
13	545	19.3	596	9	AF529819 Homo sapi
14	545	19.3	596	9	AF529820 Homo sapi
15	545	19.3	596	9	AF529821 Homo sapi
16	545	19.3	596	9	AF529822 Homo sapi
17	545	19.3	596	9	AF529824 Homo sapi
18	545	19.3	596	9	AF529825 Homo sapi
19	545	19.3	596	9	AF529826 Homo sapi
20	545	19.3	596	9	AF529827 Homo sapi
21	545	19.3	596	12	AF529830 Mus muscu
22	545	19.3	596	12	AF529831 Mus muscu
23	545	19.3	596	12	AF529833 Mus muscu
24	545	19.3	596	12	AF529835 Mus muscu
25	545	19.3	596	12	AF529837 Mus muscu
26	545	19.3	596	12	AF529839 Mus muscu
27	544	19.3	595	12	AF529834 Mus muscu
28	535	19.0	597	12	AF529841 Crictetulu
29	535	19.0	597	12	AF529843 Crictetulu
30	535	19.0	597	12	AF529844 Crictetulu
31	535	19.0	597	12	AF529845 Crictetulu
32	535	19.0	597	12	AF529848 Crictetulu
33	535	19.0	597	12	AF529849 Crictetulu
34	535	19.0	597	12	AF529850 Crictetulu
35	535	19.0	597	12	AF529851 Crictetulu
36	535	19.0	597	12	AF529852 Crictetulu
37	535	19.0	597	12	AF529853 Crictetulu
38	535	19.0	597	12	AF529854 Crictetulu
39	529	18.8	591	12	AF529855 Crictetulu
40	526	18.7	577	12	AF529840 Mus muscu
41	516	18.3	596	9	AF529818 Homo sapi
42	494	17.5	596	9	AF529817 Homo sapi
43	494	17.5	596	12	AF529832 Mus muscu
44	494	17.5	596	12	AF529836 Mus muscu
45	494	17.5	596	12	AF529838 Mus muscu
46	487	17.3	547	12	AF529856 Crictetulu
47	86	3.1	203250	9	AC007240 Homo sapi
48	79	2.8	156540	9	AC079329 Homo sapi
49	79	2.8	187836	2	AC023634 Homo sapi
50	77	2.7	41511	9	AC004603 Homo sapi
51	77	2.7	103287	2	AC084034 Homo sapi
52	77	2.7	106018	9	HS864118 Human DNA
53	77	2.7	109538	9	AC091854 Homo sapi
54	77	2.7	112804	9	AC006476 Homo sapi
55	77	2.7	121028	9	AY052369 Homo sapi
56	77	2.7	126295	9	AC079944 Homo sapi
57	77	2.7	154968	2	AC078928 Homo sapi
58	77	2.7	162958	2	AC011022 Homo sapi
59	77	2.7	177777	9	AC007342 Homo sapi
60	77	2.7	181183	2	AL391866 Homo sapi
61	77	2.7	186431	2	AC022281 Homo sapi
62	77	2.7	188863	9	AC007345 Homo sapi
63	77	2.7	192539	9	AP005264 Homo sapi
64	77	2.7	194871	9	CNS01BRV Human chr
65	77	2.7	225432	9	AF027390 Homo sapi

c 66	76	2.7	153300	2	AC011933	AC011933 Homo sapi	139	70	2.5	2460	9	AK095243	Homo sapi
c 67	76	2.7	199725	2	AC100787	AC100787 Homo sapi	140	70	2.5	22996	9	AC087175	Homo sapi
c 68	74	2.6	66452	2	AC024328	AC024328 Homo sapi	c 141	70	2.5	23379	9	AL356755	Human DNA
c 69	74	2.6	66792	9	AC008754	AC008754 Homo sapi	c 142	70	2.5	23574	6	AX326810	Sequence
c 70	74	2.6	118958	9	AC092579	AC092579 Homo sapi	c 143	70	2.5	23574	9	AF466288	Homo sapi
c 71	74	2.6	158213	2	AF214635	AF214635 Homo sapi	c 144	70	2.5	30972	9	AC123763	Homo sapi
c 72	74	2.6	158605	9	AC105081	AC105081 Homo sapi	c 145	70	2.5	35651	9	AL353798	Homo sapi
c 73	74	2.6	160371	2	AC068793	AC068793 Homo sapi	c 146	70	2.5	36921	9	AC004185	Human DNA
c 74	74	2.6	162200	9	AC012156	AC012156 Homo sapi	c 147	70	2.5	37490	9	AC026356	Homo sapi
c 75	74	2.6	167722	9	AC073548	AC073548 Homo sapi	c 148	70	2.5	38209	9	AC005542	Homo sapi
c 76	74	2.6	169234	9	AC018633	AC018633 Homo sapi	c 149	70	2.5	40709	9	L48038	Structure o
c 77	74	2.6	172816	9	AC093899	AC093899 Homo sapi	150	70	2.5	41716	9	AC011522	Homo sapi
c 78	74	2.6	176355	9	AC025518	AC025518 Homo sapi	151	70	2.5	43657	9	AC097506	Homo sapi
c 79	74	2.6	176932	9	AC016772	AC016772 Homo sapi	152	70	2.5	44336	9	AC010525	Homo sapi
c 80	74	2.6	182909	9	AC026130	AC026130 Homo sapi	153	70	2.5	62117	2	AC107891	Homo sapi
c 81	74	2.6	186870	2	AC116170	AC116170 Homo sapi	154	70	2.5	63621	9	AL590867	Homo sapi
c 82	74	2.6	187697	2	AC063929	AC063929 Homo sapi	c 155	70	2.5	64864	9	AC025260	Human DNA
c 83	74	2.6	201886	9	AC092700	AC092700 Homo sapi	c 156	70	2.5	66395	9	HS119823	Homo sapi
c 84	74	2.6	205952	9	AC024940	AC024940 Homo sapi	157	70	2.5	67550	2	AC016452	Human DNA s
c 85	74	2.6	228652	9	CNS01DWD	AL137128 Human chr	158	70	2.5	67763	2	AC103752	Homo sapi
c 86	74	2.6	250681	9	AP004282	AP004282 Homo sapi	c 159	70	2.5	67895	9	AL138807	Homo sapi
c 87	73	2.6	256319	9	AL512274	AL512274 Human DNA	160	70	2.5	68420	2	AC130285	Human DNA
c 88	73	2.6	96165	2	AC009702	AC009702 Homo sapi	c 162	70	2.5	74549	9	HSJ890015	Human DNA f
c 89	73	2.6	141175	2	AC012110	AC012110 Homo sapi	c 163	70	2.5	84132	9	AC005348	Human DNA
c 90	73	2.6	144355	2	AC016743	AC016743 Homo sapi	c 164	70	2.5	85643	2	AC015787	Homo sapi
c 91	73	2.6	153378	2	AC023536	AC023536 Homo sapi	165	70	2.5	86894	2	AL390319	Homo sapi
c 92	73	2.6	167116	9	AC009305	AC009305 Homo sapi	c 166	70	2.5	97621	9	AL607089	Homo sapi
c 93	73	2.6	176176	2	AC079360	AC079360 Homo sapi	c 167	70	2.5	97845	9	AL355337	Human DNA
c 94	73	2.6	179470	2	AL162211	AL162211 Homo sapi	c 168	70	2.5	100000	9	AP000509	Human DNA
c 95	73	2.6	183101	9	AC021915	AC021915 Homo sapi	c 169	70	2.5	106868	2	AC003023	Homo sapi
c 96	73	2.6	191652	9	AL354668	AL354668 Human DNA	c 170	70	2.5	110000	2	AC008008	Homo sapi
c 97	73	2.6	202950	9	AC017100	AC017100 Homo sapi	c 171	70	2.5	110000	2	AC080008	Continuation (2 of
c 98	73	2.6	216441	2	AP001336	AP001336 Homo sapi	c 172	70	2.5	110000	2	AC092931	Continuation (3 of
c 99	73	2.6	239008	2	AC022460	AC022460 Homo sapi	c 173	70	2.5	110000	2	AC112132	Continuation (4 of
c 100	72	2.6	89211	9	AL669821	AL669821 Human DNA	c 174	70	2.5	110000	2	AC112132	Continuation (2 of
c 101	72	2.6	104417	9	AL672220	AL672220 Homo sapi	c 175	70	2.5	110000	2	AC008053	Continuation (2 of
c 102	72	2.6	115596	9	AC096948	AC096948 Homo sapi	c 176	70	2.5	110144	9	AC008053	Continuation (2 of
c 103	72	2.6	129272	9	AL451125	AL451125 Human DNA	c 177	70	2.5	111848	2	AC124612	Homo sapi
c 104	72	2.6	143697	2	AL358572	AL358572 Homo sapi	c 178	70	2.5	113687	9	AC069281	Homo sapi
c 105	72	2.6	149827	2	AL390880	AL390880 Homo sapi	c 179	70	2.5	116236	9	AL354877	Human DNA
c 106	72	2.6	154244	2	AC073463	AC073463 Homo sapi	c 180	70	2.5	117752	9	AC006205	Homo sapi
c 107	72	2.6	156503	9	AC011511	AC011511 Homo sapi	c 181	70	2.5	118847	9	AC013413	Homo sapi
c 108	72	2.6	160903	2	AP002776	AP002776 Homo sapi	c 182	70	2.5	121856	2	AC022575	Homo sapi
c 109	72	2.6	161267	9	AC008378	AC008378 Homo sapi	c 183	70	2.5	123280	2	AC025662	Homo sapi
c 110	72	2.6	162526	2	AC024182	AC024182 Homo sapi	c 184	70	2.5	123395	9	HS889N15	Human DNA
c 111	72	2.6	162973	9	AL138958	AL138958 Human DNA	c 185	70	2.5	124102	9	AC021148	Homo sapi
c 112	72	2.6	163795	9	AP000356	AP000356 Homo sapi	c 186	70	2.5	126312	9	AC000026	Homo sapi
c 113	72	2.6	165165	9	AC095031	AC095031 Homo sapi	c 187	70	2.5	128024	9	AL590455	Human sapi
c 114	72	2.6	167386	2	AL159993	AL159993 Homo sapi	c 188	70	2.5	129413	9	HSJ709L21	Human sapi
c 115	72	2.6	171566	2	AP001885	AP001885 Homo sapi	c 189	70	2.5	132202	9	AC025218	Human sapi
c 116	72	2.6	173785	9	AP000355	AP000355 Homo sapi	c 190	70	2.5	132573	9	AC016564	Homo sapi
c 117	72	2.6	183690	9	AC091180	AC091180 Homo sapi	c 191	70	2.5	135886	9	HSX1ST2	Human Xq1
c 118	72	2.6	202943	9	AC010134	AC010134 Homo sapi	c 192	70	2.5	140606	2	AL353761	Homo sapi
c 119	72	2.6	209519	9	AC008397	AC008397 Homo sapi	c 193	70	2.5	140919	2	AC034181	Homo sapi
c 120	71	2.5	92975	9	AC002369	AC002369 Homo sapi	c 194	70	2.5	142439	9	AC008749	Homo sapi
c 121	71	2.5	110116	9	AP001931	AP001931 Homo sapi	c 195	70	2.5	145206	9	AC026170	Homo sapi
c 122	71	2.5	130642	2	AP001082	AP001082 Homo sapi	c 196	70	2.5	146360	9	HS1156N12	Human sapi
c 123	71	2.5	144588	2	AC012239	AC012239 Homo sapi	c 197	70	2.5	148179	9	AP001631	Human sapi
c 124	71	2.5	152210	9	AC007953	AC007953 Homo sapi	c 198	70	2.5	149266	2	AC021562	Homo sapi
c 125	71	2.5	153394	2	AP000727	AP000727 Homo sapi	c 199	70	2.5	149428	2	AC010264	Homo sapi
c 126	71	2.5	158349	2	AC009438	AC009438 Homo sapi	c 200	70	2.5	149870	9	AC024579	Homo sapi
c 127	71	2.5	161298	2	AP002858	AP002858 Homo sapi	c 201	70	2.5	149913	2	AC010330	Homo sapi
c 128	71	2.5	161397	2	AC090218	AC090218 Homo sapi	c 202	70	2.5	150355	9	HSJ364H10	Human sapi
c 129	71	2.5	163390	9	AL137058	AL137058 Human DNA	c 203	70	2.5	152129	9	HSJ360M22	Human sapi
c 130	71	2.5	165139	9	AC103549	AC103549 Homo sapi	c 204	70	2.5	153215	2	HS24A17	Human sapi
c 131	71	2.5	170540	2	AC092707	AC092707 Homo sapi	c 205	70	2.5	153215	2	AC023014	Homo sapi
c 132	71	2.5	189049	9	AL365495	AL365495 Human DNA	c 206	70	2.5	156422	9	AC093579	Homo sapi
c 133	71	2.5	186325	2	AC074244	AC074244 Homo sapi	c 207	70	2.5	157193	9	AL355315	Human sapi
c 134	71	2.5	193281	2	AC079376	AC079376 Homo sapi	c 208	70	2.5	157385	9	AC006023	Human sapi
c 135	71	2.5	194296	9	AC095351	AC095351 Homo sapi	c 209	70	2.5	157827	2	AC027260	Homo sapi
c 136	71	2.5	195101	9	AC110615	AC110615 Homo sapi	c 210	70	2.5	157963	9	AP002986	Homo sapi
c 137	71	2.5	199992	2	AC021522	AC021522 Homo sapi	c 211	70	2.5	158135	2	AC080040	Homo sapi
c 138	71	2.5	236723	2	AC068771	AC068771 Homo sapi							

C 212	2.5	158302	2	AC069481	Homo sapi	C 285	70	2.5	194219	2	AC023054	Homo sapi
C 213	2.5	159298	2	AL450994	Homo sapi	C 286	70	2.5	194771	2	AC109912	Homo sapi
C 214	2.5	159849	2	AC021454	Homo sapi	C 287	70	2.5	195310	2	AC024189	Homo sapi
C 215	2.5	160659	2	AC036235	Homo sapi	C 288	70	2.5	196806	9	AC104212	Homo sapi
C 216	2.5	161442	2	AL683875	Homo sapi	C 289	70	2.5	197837	9	AC093107	Homo sapi
C 217	2.5	161572	9	AC106897	Homo sapi	C 290	70	2.5	198694	9	AC012671	Homo sapi
C 218	2.5	161835	9	AC098484	Homo sapi	C 291	70	2.5	201155	9	AC009086	Homo sapi
C 219	2.5	162133	2	AC068569	Homo sapi	C 292	70	2.5	201508	2	AC026290	Homo sapi
C 220	2.5	162343	2	AC109319	Homo sapi	C 293	70	2.5	201707	2	AC131309	Homo sapi
C 221	2.5	163915	2	AC087451	Homo sapi	C 294	70	2.5	202414	2	AC090966	Papio cyn
C 222	2.5	164302	9	AC127457	Homo sapi	C 295	70	2.5	204037	2	AC015881	Homo sapi
C 223	2.5	164352	2	AC024008	Homo sapi	C 296	70	2.5	204298	9	AL662844	Human DNA
C 224	2.5	164429	9	AC026366	Homo sapi	C 297	70	2.5	206329	2	AC017113	Homo sapi
C 225	2.5	164459	9	AC092418	Homo sapi	C 298	70	2.5	206647	9	AP002898	Homo sapi
C 226	2.5	164650	2	AC093016	Homo sapi	C 299	70	2.5	207433	6	AX326809	Sequence
C 227	2.5	164838	2	AL691479	Homo sapi	C 300	70	2.5	208534	2	AC009730	Homo sapi
C 228	2.5	165447	9	AC027575	Homo sapi	C 301	70	2.5	209287	9	AC012077	Homo sapi
C 229	2.5	165499	2	AC131238	Homo sapi	C 302	70	2.5	220480	2	AC023973	Homo sapi
C 230	2.5	165861	2	AC106812	Homo sapi	C 303	70	2.5	228219	2	AC128648	Homo sapi
C 231	2.5	166867	9	AP003733	Homo sapi	C 304	70	2.5	230375	2	AC004085	Homo sapi
C 232	2.5	167343	6	AX332231	Sequence	C 305	70	2.5	236822	9	DB4394	Homo sapi
C 233	2.5	167343	6	AX335067	Sequence	C 306	70	2.5	260409	9	AC004019	Homo sapi
C 234	2.5	167343	9	HSU96629	Human chrom	C 307	70	2.5	286758	9	AC006449	Homo sapi
C 235	2.5	167357	2	AC055771	Homo sapi	C 308	70	2.5	306464	2	AC068708	Homo sapi
C 236	2.5	167631	2	AC016032	Homo sapi	C 309	70	2.5	321974	2	AC090682	Homo sapi
C 237	2.5	167668	9	AL353804	Human DNA	C 310	70	2.5	323263	2	AC079406	Homo sapi
C 238	2.5	168043	2	AC011935	Homo sapi	C 311	70	2.5	329643	2	AL512381	Homo sapi
C 239	2.5	168366	9	AC007546	Homo sapi	C 312	70	2.5	340003	9	AP001748	Homo sapi
C 240	2.5	168431	2	AC040972	Homo sapi	C 313	69	2.4	440003	9	AF317635	Homo sapi
C 241	2.5	168495	2	AC084299	Homo sapi	C 314	69	2.4	49996	9	HSJ816K9	Human DNA
C 242	2.5	168880	9	AC017104	Homo sapi	C 315	69	2.4	83120	9	HS796117	Human DNA
C 243	2.5	168887	9	AC106796	Homo sapi	C 316	69	2.4	88210	9	AL607062	Human DNA
C 244	2.5	168887	9	AL662833	Human DNA	C 317	69	2.4	126736	2	AC044788	Homo sapi
C 245	2.5	168964	2	AC131212	Homo sapi	C 318	69	2.4	129517	2	AC016400	Homo sapi
C 246	2.5	169032	9	AC016931	Homo sapi	C 319	69	2.4	147999	2	AL161635	Homo sapi
C 247	2.5	169118	2	AC096645	Homo sapi	C 320	69	2.4	148295	9	AC090510	Homo sapi
C 248	2.5	170444	2	AL359977	Homo sapi	C 321	69	2.4	154037	9	AC072054	Homo sapi
C 249	2.5	170569	2	AC108109	Homo sapi	C 322	69	2.4	160815	2	AP001895	Homo sapi
C 250	2.5	170882	2	AC023429	Homo sapi	C 323	69	2.4	162191	9	AC109458	Homo sapi
C 251	2.5	170970	9	AL354707	Human DNA	C 324	69	2.4	163372	2	AC087744	Homo sapi
C 252	2.5	171829	9	AC022116	Homo sapi	C 325	69	2.4	163569	2	AC023293	Homo sapi
C 253	2.5	172049	2	AC008414	Homo sapi	C 326	69	2.4	165373	9	AC067957	Homo sapi
C 254	2.5	173029	9	AC002059	Homo sapi	C 327	69	2.4	172525	9	AC008514	Homo sapi
C 255	2.5	173085	2	AC127524	Homo sapi	C 328	69	2.4	185075	2	AC118583	Papio cyn
C 256	2.5	174058	2	AC068695	Homo sapi	C 329	69	2.4	186755	9	AC024022	Homo sapi
C 257	2.5	174231	9	AC021054	Homo sapi	C 330	69	2.4	186783	9	AC090355	Homo sapi
C 258	2.5	175072	2	AC022358	Homo sapi	C 331	69	2.4	186902	9	AC010863	Homo sapi
C 259	2.5	175625	2	AC068475	Homo sapi	C 332	69	2.4	187167	2	AP001563	Homo sapi
C 260	2.5	175980	2	AC025859	Homo sapi	C 333	69	2.4	191436	9	AL158830	Human DNA
C 261	2.5	176206	9	AC009967	Homo sapi	C 334	69	2.4	228652	9	CNS01DWD	Human chr
C 262	2.5	179441	2	AC044879	Homo sapi	C 335	68	2.4	52350	9	AC092330	Homo sapi
C 263	2.5	179843	2	AC131159	Homo sapi	C 336	68	2.4	69901	9	HS694E4	Human DNA
C 264	2.5	180392	9	AC025945	Homo sapi	C 337	68	2.4	87402	9	HSJ437M21	Human DNA
C 265	2.5	181923	9	AC022872	Homo sapi	C 338	68	2.4	94568	9	AC114973	Homo sapi
C 266	2.5	182408	2	AL358772	Homo sapi	C 339	68	2.4	100072	9	AL137144	Human DNA
C 267	2.5	184268	2	AC011401	Homo sapi	C 340	68	2.4	108893	9	AC079169	Homo sapi
C 268	2.5	184362	2	AL450268	Homo sapi	C 341	68	2.4	114023	2	HSJ16B5	Homo sapi
C 269	2.5	184470	9	AL589823	Human DNA	C 342	68	2.4	125055	2	HS167F1	Human sapien
C 270	2.5	184825	9	AL391380	Human DNA	C 343	68	2.4	126152	9	HS433M19	s
C 271	2.5	185574	9	AC073341	Homo sapi	C 344	68	2.4	129338	9	AC016597	Homo sapi
C 272	2.5	186135	2	AC124947	Homo sapi	C 345	68	2.4	130824	9	AC078880	Homo sapi
C 273	2.5	186558	2	AC079031	Homo sapi	C 346	68	2.4	158143	2	AC016893	Homo sapi
C 274	2.5	186780	9	AC005740	Homo sapi	C 347	68	2.4	162124	2	AC113191	Homo sapi
C 275	2.5	188049	2	AC069391	Homo sapi	C 348	68	2.4	164076	9	AC021059	Homo sapi
C 276	2.5	188141	2	AC023120	Homo sapi	C 349	68	2.4	165287	2	AC034165	Homo sapi
C 277	2.5	188481	9	AL138836	Human DNA	C 350	68	2.4	175670	2	AC103953	Homo sapi
C 278	2.5	188645	9	AC012047	Homo sapi	C 351	68	2.4	178336	9	AC073359	Homo sapi
C 279	2.5	188823	2	AC016990	Homo sapi	C 352	68	2.4	185551	9	HSJ17K16	Human DNA
C 280	2.5	188846	9	AC009030	Homo sapi	C 353	68	2.4	187469	2	AC058793	Homo sapi
C 281	2.5	189156	9	AC008669	Homo sapi	C 354	68	2.4	188948	9	AC005832	Homo sapi
C 282	2.5	189972	2	AC107214	Homo sapi	C 355	68	2.4	197652	9	AC012074	Homo sapi
C 283	2.5	191872	2	AC129837	Papio cyn	C 356	68	2.4	201973	2	AC087314	Homo sapi
C 284	2.5	194197	9	AC064871	Homo sapi	C 357	68	2.4	224044	2	AC010882	Homo sapi

C 358	67	2.4	363	11	GL3005	GL3005 SWSS2845 Er	C 431	66	2.3	147201	2	AC053469	AC053469 Homo sapi
C 359	67	2.4	401	9	HSU41198	U41198 Human (TTTG	432	66	2.3	148046	9	AC034226	AC034226 Homo sapi
C 360	67	2.4	4565	9	AF187983	AF187983 Homo sapi	433	66	2.3	148911	2	AL353142	AL353142 Homo sapi
C 361	67	2.4	5055	2	AC005843	AC005843 Homo sapi	434	66	2.3	149125	2	AC069401	AC069401 Homo sapi
C 362	67	2.4	30960	2	AC003102	AC003102 Homo sapi	435	66	2.3	150791	9	AL162458	AL162458 Human DNA
C 363	67	2.4	34164	2	AC090913	AC090913 Homo sapi	C 436	66	2.3	151600	2	AC010278	AC010278 Homo sapi
C 364	67	2.4	51407	9	AL357556	AL357556 Human DNA	437	66	2.3	152709	2	AC114799	AC114799 Homo sapi
C 365	67	2.4	52459	9	AF228703	AF228703 Homo sapi	C 438	66	2.3	152794	9	AL359085	AL359085 Human DNA
C 366	67	2.4	61294	9	AF671934	AF671934 Human DNA	439	66	2.3	155067	9	AC010301	AC010301 Homo sapi
C 367	67	2.4	64167	9	HSBA8982	AL080245 Human DNA	440	66	2.3	155470	9	AC104448	AC104448 Homo sapi
C 368	67	2.4	67131	2	AC103791	AC103791 Homo sapi	C 441	66	2.3	157377	2	AC022827	AC022827 Homo sapi
C 369	67	2.4	82700	2	AC015635	AC015635 Homo sapi	C 442	66	2.3	159007	9	AC020658	AC020658 Homo sapi
C 370	67	2.4	82700	2	AC015635	AC015635 Homo sapi	C 443	66	2.3	160161	2	AC007846	AC007846 Homo sapi
C 371	67	2.4	93385	2	AC109513	AC109513 Homo sapi	C 444	66	2.3	160417	2	AC024317	AC024317 Homo sapi
C 372	67	2.4	96002	2	AC109326	AC109326 Homo sapi	445	66	2.3	163540	2	AC024229	AC024229 Homo sapi
C 373	67	2.4	101077	2	AC024051	AC024051 Homo sapi	446	66	2.3	163197	2	AC091392	AC091392 Homo sapi
C 374	67	2.4	104228	9	AL354751	AL354751 Human DNA	C 447	66	2.3	164500	2	AC069171	AC069171 Homo sapi
C 375	67	2.4	118234	9	AL355476	AL355476 Human DNA	C 448	66	2.3	165496	9	AC104461	AC104461 Homo sapi
C 376	67	2.4	124001	9	HSB86K2	AL031295 Human DNA	C 449	66	2.3	167254	9	CNS05TDS	AL357093 Human chr
C 377	67	2.4	124814	9	AL161436	AL161436 Human DNA	C 450	66	2.3	167390	2	AC007263	AC007263 Homo sapi
C 378	67	2.4	137830	9	AC003029	AC003029 Homo sapi	451	66	2.3	168751	2	AC068969	AC068969 Homo sapi
C 379	67	2.4	142060	2	AL353642	AL353642 Homo sapi	452	66	2.3	168956	2	AC090286	AC090286 Homo sapi
C 380	67	2.4	147888	2	AC016538	AC016538 Homo sapi	C 453	66	2.3	170883	9	AC067805	AC067805 Homo sapi
C 381	67	2.4	149266	2	AC116022	AC116022 Homo sapi	C 454	66	2.3	170891	2	AC002118	AC002118 Homo sapi
C 382	67	2.4	153875	9	AC003682	AC003682 Homo sapi	C 455	66	2.3	171452	9	AP000476	AP000476 Homo sapi
C 383	67	2.4	156997	9	AC011465	AC011465 Homo sapi	C 456	66	2.3	172654	2	AC067737	AC067737 Homo sapi
C 384	67	2.4	159322	2	AC026085	AC026085 Homo sapi	C 457	66	2.3	172867	2	AC026207	AC026207 Homo sapi
C 385	67	2.4	161788	9	AC009314	AC009314 Homo sapi	C 458	66	2.3	173226	9	CNS01DTQ	AL132838 Human chr
C 386	67	2.4	162728	9	AP002906	AP002906 Homo sapi	C 459	66	2.3	175652	9	AC084864	AC084864 Homo sapi
C 387	67	2.4	164034	9	AC011489	AC011489 Homo sapi	C 460	66	2.3	176715	2	AC073474	AC073474 Homo sapi
C 388	67	2.4	167072	2	AC060780	AC060780 Homo sapi	461	66	2.3	176861	9	AC097065	AC097065 Homo sapi
C 389	67	2.4	170595	9	AC016334	AC016334 Homo sapi	C 462	66	2.3	177819	9	AC100771	AC100771 Homo sapi
C 390	67	2.4	176582	9	AC016334	AC016334 Homo sapi	463	66	2.3	179462	2	AC021102	AC021102 Homo sapi
C 391	67	2.4	176121	2	AC013489	AC003010 Homo sapi	464	66	2.3	180736	2	AC023925	AC023925 Homo sapi
C 392	67	2.4	176709	9	AC011737	AC011737 Homo sapi	C 465	66	2.3	182288	9	AC018362	AC018362 Homo sapi
C 393	67	2.4	176973	2	AC130449	AC130449 Homo sapi	466	66	2.3	188050	2	AL391845	AL391845 Homo sapi
C 394	67	2.4	180774	2	AC061979	AC061979 Homo sapi	467	66	2.3	190619	9	AC108172	AC108172 Homo sapi
C 395	67	2.4	182123	2	AC002399	AC002399 Homo sapi	C 468	66	2.3	191426	2	AC026086	AC026086 Homo sapi
C 396	67	2.4	183599	2	AC073501	AC073501 Homo sapi	C 469	66	2.3	193281	2	AC079376	AC079376 Homo sapi
C 397	67	2.4	188846	9	AC009030	AC009030 Homo sapi	470	66	2.3	194545	2	AC087482	AC087482 Homo sapi
C 398	67	2.4	189271	9	AL138752	AL138752 Human DNA	471	66	2.3	198017	9	AC012314	AC012314 Homo sapi
C 399	67	2.4	195648	2	AF215848	AF215848 Homo sapi	C 472	66	2.3	198316	2	AC019157	AC019157 Homo sapi
C 400	67	2.4	199451	2	AC093572	AC093572 Homo sapi	C 473	66	2.3	200430	9	AC011500	AC011500 Homo sapi
C 401	67	2.4	202638	2	AC103959	AC103959 Homo sapi	C 474	66	2.3	205307	2	CNS01IRG3	AL157736 Homo sapi
C 402	67	2.4	206596	2	AC087624	AC087624 Homo sapi	475	66	2.3	205307	2	AC009968	AC009968 Homo sapi
C 403	67	2.4	208612	2	AC108668	AC108668 Homo sapi	476	66	2.3	208396	9	AC003101	AC003101 Homo sapi
C 404	67	2.4	214530	9	AC020916	AC020916 Homo sapi	C 477	66	2.3	209752	2	AC079620	AC079620 Homo sapi
C 405	67	2.4	237931	9	AC022098	AC022098 Homo sapi	C 478	66	2.3	212030	2	AC019148	AC019148 Homo sapi
C 406	67	2.4	242970	2	AC004803	AC004803 Homo sapi	C 479	66	2.3	213865	9	AC093166	AC093166 Homo sapi
C 407	67	2.4	288888	9	HSB310932	ALJ310932 Homo sapi	480	66	2.3	214017	2	AC026359	AC026359 Homo sapi
C 408	67	2.4	298575	2	HSBA2284	AL080244 Homo sapi	C 481	66	2.3	214162	9	AC084865	AC084865 Homo sapi
C 409	66	2.3	2405	9	AK094131	AK094131 Homo sapi	C 482	66	2.3	220807	9	AC093334	AC093334 Homo sapi
C 410	66	2.3	2847	9	AK094891	AK094891 Homo sapi	483	66	2.3	340000	9	AP001690	AP001690 Homo sapi
C 411	66	2.3	34639	9	AC005765	AC005765 Homo sapi	484	65	2.3	2687	9	AK026456	AK026456 Homo sapi
C 412	66	2.3	53196	9	AC108047	AC108047 Homo sapi	C 485	65	2.3	28136	9	HSL184D6	Z49236 Human DNA s
C 413	66	2.3	62449	2	AC061985	AC061985 Homo sapi	486	65	2.3	67187	9	AC108744	AC108744 Homo sapi
C 414	66	2.3	78036	9	AC005755	AC005755 Homo sapi	487	65	2.3	98387	9	AC003077	AC003077 Human IAC
C 415	66	2.3	89250	2	AC016334	AC016334 Homo sapi	488	65	2.3	102477	9	AC008935	AC008935 Homo sapi
C 416	66	2.3	90876	9	AC015910	AC015910 Homo sapi	C 489	65	2.3	107919	2	AL391274	AL391274 Homo sapi
C 417	66	2.3	92975	9	AC002369	AC002369 Homo sapi	490	65	2.3	137496	9	AC005288	AC005288 Homo sapi
C 418	66	2.3	109445	9	AC008946	AC008946 Homo sapi	C 491	65	2.3	146868	2	AC016802	AC016802 Homo sapi
C 419	66	2.3	111107	9	AC005924	AC005924 Homo sapi	492	65	2.3	148120	9	AC090051	AC090051 Homo sapi
C 420	66	2.3	111682	9	AC005162	AC005162 Homo sapi	C 493	65	2.3	153148	2	AC017030	AC017030 Homo sapi
C 421	66	2.3	112573	9	AL353151	AL353151 Human DNA	C 494	65	2.3	154968	9	AL596342	AL596342 Human DNA
C 422	66	2.3	123149	9	AC008655	AC008655 Homo sapi	C 495	65	2.3	160615	9	AC112236	AC112236 Homo sapi
C 423	66	2.3	124310	2	AC008587	AC008587 Homo sapi	C 496	65	2.3	160903	2	AP002776	AP002776 Homo sapi
C 424	66	2.3	124949	9	AL358794	AL358794 Human DNA	497	65	2.3	162294	2	AC112321	AC112321 Homo sapi
C 425	66	2.3	130609	2	AC023217	AC023217 Homo sapi	C 498	65	2.3	166558	2	AC112141	AC112141 Homo sapi
C 426	66	2.3	130853	9	AC091974	AC091974 Homo sapi	C 499	65	2.3	169339	2	AC110008	AC110008 Homo sapi
C 427	66	2.3	132644	9	AC104801	AC104801 Homo sapi	C 500	65	2.3	169892	9	AC068397	AC068397 Homo sapi
C 428	66	2.3	134462	9	AC001231	AC001231 Genomic s							
C 429	66	2.3	142698	9	AC025429	AC025429 Homo sapi							
C 430	66	2.3	147140	9	AC010226	AC010226 Homo sapi							

[illegible]

Db	1261	AGCATGTTTATGTTTGTACAAAAGAAGATTGTTATGGTGGGATGGAGGTATAGAC	1320
QY	1324	CATGATGTCACCTTCAAGCTACTTTAATAAAGATCTTAAATGGGCGAGGACACTGT	1383
Db	1321	CATGATGTCACCTTCAAGCTACTTTAATAAAGATCTTAAATGGGCGAGGACACTGT	1380
QY	1384	GAACAAGACACCTTAATAATGGTGTGATGTCGAAGTAGCAAAATCTTCTGGAACGCCAA	1443
Db	1381	GAACAAGACACCTTAATAATGGTGTGATGTCGAAGTAGCAAAATCTTCTGGAACGCCAA	1440
QY	1444	CTCTTTAAGGAAGTCCCTAATTTAGAACACCCCAAACTTCACATATCATTAATGACA	1503
Db	1441	CTCTTTAAGGAAGTCCCTAATTTAGAACACCCCAAACTTCACATATCATTAATGACA	1500
QY	1504	AACAAATGGAAGAAAGTTGCTTGAATGTTGGGAGAGGAAATCTATTGGCTCTCGTGGG	1563
Db	1501	AACAAATGGAAGAAAGTTGCTTGAATGTTGGGAGAGGAAATCTATTGGCTCTCGTGGG	1560
QY	1564	TCCTTCATCTCAGAAATGCCAATCAGGTCGAAGTTTGCTACATTTTGATGTGTGAT	1623
Db	1561	TCCTTCATCTCAGAAATGCCAATCAGGTCGAAGTTTGCTACATTTTGATGTGTGAT	1620
QY	1624	GCTTCTCCAAAAGGTATTTAACTATATAAGAGCTTGTGACAAAACAATGATAAGC	1683
Db	1621	GCTTCTCCAAAAGGTATTTAACTATATAAGAGCTTGTGACAAAACAATGATAAGC	1680
QY	1684	TCCGAACCGTGGCACACGCTCATAGTCTAGCTGCTGGGAGTTGAGGAGGAGATGG	1743
Db	1681	TCCGAACCGTGGCACACGCTCATAGTCTAGCTGCTGGGAGTTGAGGAGGAGATGG	1740
QY	1744	CTTGAACACAGTGTTCAGGCCAGCCTGGGCAACATACCAAGATCCCTGCTCTCAAAA	1803
Db	1741	CTTGAACACAGTGTTCAGGCCAGCCTGGGCAACATACCAAGATCCCTGCTCTCAAAA	1800
QY	1804	AAAAAAGGAGAGAGAGGCGCGGTGTGTCTCAGCGCTCTAATCCAG	1863
Db	1801	AAAAAAGGAGAGAGAGGCGCGGTGTGTCTCAGCGCTCTAATCCAG	1860
QY	1864	CACTTGGAGCGCGAGCGCGGATCACCTGTGGTCAGAGTTTGACACCGCTGGC	1923
Db	1861	CACTTGGAGCGCGAGCGCGGATCACCTGTGGTCAGAGTTTGACACCGCTGGC	1920
QY	1924	CAACATGCAAAACCCCTGTCTACTCAAAATGCAAAATAGCCAGCGTGTAGCAGG	1983
Db	1921	CAACATGCAAAACCCCTGTCTACTCAAAATGCAAAATAGCCAGCGTGTAGCAGG	1980
QY	1984	CACCTGTATCCAGCTACTTTGGAGGCTGAGGAGGAGATCCCTTGAACCCAGGAGT	2043
Db	1981	CACCTGTATCCAGCTACTTTGGAGGCTGAGGAGGAGATCCCTTGAACCCAGGAGT	2040
QY	2044	GGAGTTGCATGAGATCGTGGCTGTCACCTCCAGCTCCAGCTGGGCGACAGCAAGA	2103
Db	2041	GGAGTTGCATGAGATCGTGGCTGTCACCTCCAGCTGGGCGACAGCAAGA	2100
QY	2104	CTCTGTCTCAGAAAAAAG	2163
Db	2101	CTCTGTCTCAGAAAAAAG	2160
QY	2164	GAGAGGATGGGAAGCATTCAGGAATTTGCTTTATCCAAACAAATGTAAGGAGCC	2223
Db	2161	GAGAGGATGGGAAGCATTCAGGAATTTGCTTTATCCAAACAAATGTAAGGAGCC	2220
QY	2224	AATAAGGATCCCTATTGTCTCTTTTGGTGTCTATTGTCCTTAACTGCTTTGAC	2283
Db	2221	AATAAGGATCCCTATTGTCTCTTTTGGTGTCTATTGTCCTTAACTGCTTTGAC	2280
QY	2284	AGTGAGAAAAATATTCAGAAATACCATATCCCTGTGCGGTATTATACCTAGCAACCTTGC	2343
Db	2281	AGTGAGAAAAATATTCAGAAATACCATATCCCTGTGCGGTATTATACCTAGCAACCTTGC	2340
QY	2344	AATCAAGATGAGCAGATCCACAGAGAACTTGAATGCACAACTGCTTTAATCTTA	2403
Db	2341	AATCAAGATGAGCAGATCCACAGAGAACTTGAATGCACAACTGCTTTAATCTTA	2400
QY	2404	TTGTACATAAGCTTTCTAAAAAGAGTTAAAAATTTGTTACTTTCATTTATATTTTAA	2463
Db	2401	TTGTACATAAGCTTTCTAAAAAGAGTTAAAAATTTGTTACTTTCATTTATATTTTAA	2460
QY	2464	TATTATTTTGGCTCTAATGATTTTTTATTAAACATGATTCTCTTTCTGATATATTGAAT	2523
Db	2461	TATTATTTTGGCTCTAATGATTTTTTATTAAACATGATTCTCTTTCTGATATATTGAAT	2520
QY	2524	GGAGTCTCAAAAGCTTCATAAATTTTAACTTTTAAAGATGATCTTAATAACACCTATGTA	2583
Db	2521	GGAGTCTCAAAAGCTTCATAAATTTTAACTTTTAAAGATGATCTTAATAACACCTATGTA	2580
QY	2584	ATTGTAAACATTTGCAGTAAATGCTACGAAGCCATTTCTCTTGATTTTACTAAACCTTT	2643
Db	2581	ATTGTAAACATTTGCAGTAAATGCTACGAAGCCATTTCTCTTGATTTTACTAAACCTTT	2640
QY	2644	ATGACAGCAAAATTTGCTTCTGGCTCACCTTTCAATCAGTTAAATAAATGATAAATTTT	2703
Db	2641	ATGACAGCAAAATTTGCTTCTGGCTCACCTTTCAATCAGTTAAATAAATGATAAATTTT	2700
QY	2704	GGAAGCTCTGAAGATAAAATACCAAAATAAATAATATAAAAGTGATTTATATGAAGTTAA	2763
Db	2701	GGAAGCTCTGAAGATAAAATACCAAAATAAATAATATAAAAGTGATTTATATGAAGTTAA	2760
QY	2764	AATAAAAAATCAGTATGATGAATAAACTTG	2794
Db	2761	AATAAAAAATCAGTATGATGAATAAACTTG	2791
RESULT 2	AB040430	11204 bp DNA linear	PRI 03-OCT-2000
LOCUS	AB040430	Homo sapiens AID gene for activation-induced cytidine deaminase, complete cds.	
ACCESSION	AB040430		
VERSION	AB040430.1	GI:9988407	
KEYWORDS	AID; activation-induced cytidine deaminase.		
SOURCE	Homo sapiens DNA.		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	Muto, T., Muramatsu, M., Tanikawa, M., Kinoshita, K. and Honjo, T.		
TITLE	Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene		
JOURNAL	Genomics 68 (1), 85-88 (2000)		
MEDLINE	20408890		
REFERENCE			
AUTHORS	Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Bufourcq-Lagelouse, R., Genin, A., Tezcan, I., Ersoy, F., Kayserli, H., Ugazio, A.G., Brousse, N., Muramatsu, M., Notarangelo, L.D., Kinoshita, K., Honjo, T., Fischer, A. and Durandy, A.		
TITLE	Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)		
JOURNAL	Cell 102 (5), 565-575 (2000)		
MEDLINE	20460541		
REFERENCE			
AUTHORS	Muto, T., Muramatsu, M., Tanikawa, M., Kinoshita, K. and Honjo, T.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan (E-mail: honjo@four.med.kyoto-u.ac.jp, Tel: 81-75-753-4371(ex.4371), Fax: 81-75-753-4388)		
FEATURES			
source	1..11204	Location/Qualifiers	
gene	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	Join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)		
	/gene="AID"		
CDS	Join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)		
	/gene="AID"		

Db	9795	TAATTTAGAAACACCCACAACCTTCACATATCATATAATTACCAACAACATTTGGAGGAAGTT	98954
Qy	1522	CGTTTGAATGTTGGGAGAGAGAAATCATTTGGGCTCGCTGGGTCCTTTCATCTCAGAAAT	1581
Db	9855	CGTTTGAATGTTGGGAGAGAGAAATCATTTGGGCTCGCTGGGTCCTTTCATCTCAGAAAT	9914
Qy	1582	GCCAATCAGGTCGAAGGTTTCCTACATTTTCATGTGTGTATGCTCTCCCAAGGATA	1641
Db	9915	GCCAATCAGGTCGAAGGTTTCCTACATTTTCATGTGTGTATGCTCTCCCAAGGATA	9974

Qy	1942	TTACTATATAAGGAGTGTGGTCTCCTCAAAACCAAGATGATAAAGCTCCGACACCG
Db	9975	TTACTATATAAGGAGTGTGGTCAAAACAAGAATGATAAAGCTCGCAACCGTGCCACACG
Qy	1702	CTCATAGTTCTAGCTGCTTGGGAGTTTGAGGAGGAGGATGGCTTGAACACAGGTGTTC
Db	10035	CTCATAGTTCTAGCTGCTTGGGAGTTTGAGGAGGAGGATGGCTTGAACACAGGTGTTC
Qy	1762	AGGCCAGCCTGGGCAACATTAACAGATCTCTCTCAAAAAAAAAGAAAGAA
Db	10095	AGGCCAGCCTGGGCAACATTAACAGATCTCTCTCAAAAAAAAAGAAAGAAAGAA
Qy	1822	AGAGAGAGGC GGCGCTGTGCTCACGCTGTGAATCCAGCACCTTTGGAGGCGCAGC
Db	10155	AGAGAGAGGC GGCGCTGTGCTCACGCTGTGAATCCAGCACCTTTGGAGGCGCAGC
Qy	1882	CGGCGGATCACCTGTGGTCAGGAGTTTGAGACCAGCTTGCCCAACATGCAAAACCCCG
Db	10215	CGGCGGATCACCTGTGGTCAGGAGTTTGAGACCAGCTTGCCCAACATGCAAAACCCCG
Qy	1942	TCTCTACTCAAAATGCAAAAAATTAGCCAGCGGTGTAGCAGGCACCTGTATCCCAGCTA
Db	10275	TCTCTACTCAAAATGCAAAAAATTAGCCAGCGGTGTAGCAGGCACCTGTATCCCAGCTA
Qy	2002	CTTGGGAGGCTGAGGAGAGAAATCGCTTGAACCCAGGAGGTGGAGGTTGCAGTAAGCTG
Db	10335	CTTGGGAGGCTGAGGAGAGAAATCGCTTGAACCCAGGAGGTGGAGGTTGCAGTAAGCTG
Qy	2062	AGATCGTGGCTTGCATCTCAGCGCTGGGCGACAAAGACAAGCTCTCTCAGAAAAAA
Db	10395	AGATCGTGGCTTGCATCTCAGCGCTGGGCGACAAAGACAAGCTCTCTCAGAAAAAA
Qy	2122	AAAAAAAAGAGAGAGAGAGAAAGAGAACAAATATTTGGGAGAGAAGGATGGGGAAGCA
Db	10455	AAAAAAAAGAGAGAGAGAGAAAGAGAACAAATATTTGGGAGAGAAGGATGGGGAAGCA
Qy	2182	TTGCAAGGAAATTTGCTTTTTATCCAACAAAATGTAAGGAGCCATAAAGGATCCCTATTT
Db	10515	TTGCAAGGAAATTTGCTTTTTATCCAACAAAATGTAAGGAGCCATAAAGGATCCCTATTT
Qy	2242	GTCCTCTTTTGGTGCTATTTTGTCCCTAACCAACTGCTTTGCACAGTGAGAAAAATATTCAG
Db	10575	GTCCTCTTTTGGTGCTATTTTGTCCCTAACCAACTGCTTTGCACAGTGAGAAAAATATTCAG
Qy	2302	AATAACCATATCCCTGTGGCGTTATTAACCTAGCAACCCCTTGCAATGAAGATGAGCAGATC
Db	10635	AATAACCATATCCCTGTGGCGTTATTAACCTAGCAACCCCTTGCAATGAAGATGAGCAGATC
Qy	2362	CACAGGAAACCTTGAATGCACAACTGCTTTATTTTAACTTTATTTGATACATAAGTTGTAA
Db	10695	CACAGGAAACCTTGAATGCACAACTGCTTTATTTTAACTTTATTTGATACATAAGTTGTAA
Qy	2422	AAGAGTTAAAAATTTACTTCATGTAATTCATTTATATTTATTTTGGCTCTTAAT
Db	10755	AAGAGTTAAAAATTTACTTCATGTAATTCATTTATATTTATTTTGGCTCTTAAT
Qy	2482	GATTTTTTATTAACATGATTTCCCTTTTCTGATATATTTGAAATGGAGTCTCAAAGCTTCAT
Db	10815	GATTTTTTATTAACATGATTTCCCTTTTCTGATATATTTGAAATGGAGTCTCAAAGCTTCAT
Qy	2542	AAATTTTAACTTTAGAAATGATTTCTTAATAACACAGTATGTAATTTGTAACATTCAGTAA
Db	10875	AAATTTTAACTTTAGAAATGATTTCTTAATAACACAGTATGTAATTTGTAACATTCAGTAA

QY	2602	TGCTGTACGAGCCATTTCTCTGTGATTTTATAGTAACTTTTATGACGCAAAATTTGCTT	2661
Db	10935	TGCTGTACGAGCCATTTCTCTGTGATTTTATAGTAACTTTTATGACGCAAAATTTGCTT	10994
QY	2662	CTGGCTCAGTTTCAATCAGTAAATTAATGATAAATTTTGGAGCGTGTGAAGATAA	2721
Db	10995	CTGGCTCAGTTTCAATCAGTAAATTAATGATAAATTTTGGAGCGTGTGAAGATAA	11054
QY	2722	ATACCAATAAATAATATAAAGTGTATATGAGTTAAATTAATAAATAAATCAGTATGA	2781
Db	11055	ATACCAATAAATAATATAAAGTGTATATGAGTTAAATTAATAAATAAATCAGTATGA	11114
QY	2782	TGGAATAAACTTGA	2795
Db	11115	TGGAATAAACTTGA	11128

AC092184 71132 bp DNA linear PRI 12-JUN-2002
Homo sapiens 12 BAC RP11-438L7 (Roswell Park Cancer Institute Human BAC Library) complete sequence.

AC092184 AC013443
AC092184.7 GI:21206067
HTG.
human.

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 71132)

REFERENCE
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C., Alsbrooks,S.L., Anarstange,H.C., Are,J.R., Ayale,M., Banks,T., Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falis,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shooshkar,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Zorrilla,S., Kucherlapati,R., Weinstein,G. and Gibbs.R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Direct Submission
Unpublished
2 (bases 1 to 71132)
Worley,K.C.
Direct Submission
Submitted (25-JUN-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 71132)
Worley,K.C.
Direct Submission
Submitted (18-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 71132)
Worley,K.C.
Direct Submission
Submitted (25-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 71132)
Worley,K.C.
Direct Submission
Submitted (12-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 25, 2002 this sequence version replaced gi:20901754.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-14) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES	Location/Qualifiers
source	1. 71132 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="12" /clone="RP11-438L7" /complement(1..1448) /note="overlaps bases 1..1448 of clone AC012490"
misc_feature	

Qy	1642	TTAACTATATAGAGAGTTGTGACAAAACAGAAATGATAAAGCTGCGAACCCGTGGCACAGC	1701
Db	44518	TTAACTATATAGAGAGTTGTGACAAAACAGAAATGATAAAGCTGCGAACCCGTGGCACAGC	44577
Qy	1702	CTCATAGTTCTAGCTGCTTGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGTGTTC	1761
Db	44578	CTCATAGTTCTAGCTGCTTGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGTGTTC	44637
Qy	1762	AGCCAGCCTGGGCAACATACACAGCTCTCTCAAAAAAAGAAAAAAGAA	1821
Db	44638	AGCCAGCCTGGGCAACATACACAGCTCTCTCAAAAAAAGAAAAAAGAA	44697
Qy	1822	AGAGAGAGGCGGCGCTGTGGCTCAGCCGTGAATCCAGCACTTTGGGAGCGCGAGC	1881
Db	44698	AGAGAGAGGCGGCGCTGTGGCTCAGCCGTGAATCCAGCACTTTGGGAGCGCGAGC	44757
Qy	1882	CGGCGGATCACCCTGTGTGAGGAGTTGAGACAGCGCTGGCCCAACATGGCAAAACCCCG	1941
Db	44758	CGGCGGATCACCCTGTGTGAGGAGTTGAGACAGCGCTGGCCCAACATGGCAAAACCCCG	44817
Qy	1942	TCTGTACTCAAAATGCAAAATAGCCAGGCGTGTAGCAGCGCACCTGTATCCAGCTA	2001
Db	44818	TCTGTACTCAAAATGCAAAATAGCCAGGCGTGTAGCAGCGCACCTGTATCCAGCTA	44877
Qy	2002	CTTGGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGAGTGGAGTTCAGTAAGCTG	2061
Db	44878	CTTGGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGAGTGGAGTTCAGTAAGCTG	44937
Qy	2062	AGATCGTCCCTTGCATCCAGCTGGCGCAAGAGCAAGACATCTGTCTCAGAAAAAA	2121
Db	44938	AGATCGTCCCTTGCATCCAGCTGGCGCAAGAGCAAGACATCTGTCTCAGAAAAAA	44997
Qy	2122	AAAAAAAAGAGAGAGAGAGAGAAATATTTGGGAGAGAGGATGGGCAAGCA	2181
Db	44998	AAAAAAAAGAGAGAGAGAGAGAAATATTTGGGAGAGAGGATGGGCAAGCA	45057
Qy	2182	TTGCAAGAAATGTGCTTATCCACAAATGTAAAGAGCCAAATAGGATCCCTATTT	2241
Db	45058	TTGCAAGAAATGTGCTTATCCACAAATGTAAAGAGCCAAATAGGATCCCTATTT	45117
Qy	2242	GTCTCTTTTGTGTCTATTTGTCCTTCAACACTGCTTTCACAGTGAGAAAAATATTCAG	2301
Db	45118	GTCTCTTTTGTGTCTATTTGTCCTTCAACACTGCTTTCACAGTGAGAAAAATATTCAG	45177
Qy	2302	AATAACCATATCCCTGTGCCGTATTACCTAGCAACCCCTTGCAATGAAGATGACAGATC	2361
Db	45178	AATAACCATATCCCTGTGCCGTATTACCTAGCAACCCCTTGCAATGAAGATGACAGATC	45237
Qy	2362	CACAGGAAACTTGAATGCACACTGCTTATTTTATCTATTGTACATAAGTTGTGAA	2421
Db	45238	CACAGGAAACTTGAATGCACACTGCTTATTTTATCTATTGTACATAAGTTGTGAA	45297
Qy	2422	AAGAGTTAAAAATTTTACTTTCATGTATTCATTTATATTTTATTTTGGTCTCAAT	2481
Db	45298	AAGAGTTAAAAATTTTACTTTCATGTATTCATTTATATTTTATTTTGGTCTCAAT	45357
Qy	2482	GATTTTTTATTAACATGATTTTCTTTTCTGATATATTGAATGGAGTCTCAAAAGCTTCAT	2541
Db	45358	GATTTTTTATTAACATGATTTTCTTTTCTGATATATTGAATGGAGTCTCAAAAGCTTCAT	45417
Qy	2542	AAATTTTATACCTTTAGAAATGATCTAATACACAGTATGTAATGTGAACATTGCAGTAA	2601
Db	45418	AAATTTTATACCTTTAGAAATGATCTAATACACAGTATGTAATGTGAACATTGCAGTAA	45477
Qy	2602	TGGTGTACGAAGCATTTCTCTGTATTTTATTAATCTTTTATGACAGCAAAATTCCTT	2661
Db	45478	TGGTGTACGAAGCATTTCTCTGTATTTTATTAATCTTTTATGACAGCAAAATTCCTT	45537
Qy	2662	CTGCTCAGCTTTCAATCAGTTAAATAAATGATAAATTTTGAAGCTCTGAAGATAAA	2721
Db	45538	CTGCTCAGCTTTCAATCAGTTAAATAAATGATAAATTTTGAAGCTCTGAAGATAAA	45597

Qy	2722	ATACCAATAAATAATATAAAGCTGATTTATATGAAGTTTAAATAAAAAATCAGTATGA	2781
Db	45598	ATACCAATAAATAATATAAAGCTGATTTATATGAAGTTTAAATAAAAAATCAGTATGA	45657
Qy	2782	TGGAATAAACTTGA	2795
Db	45658	TGGAATAAACTTGA	45671
RESULT 4			
BC006296			
LOCUS			
DEFINITION	Homo sapiens, activation-induced cytidine deaminase, clone	1837 bp	linear PRI 12-JUL-2001
ACCESSION	BC006296		
VERSION	BC006296.1	GI:13623400	
KEYWORDS	MGC.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
JOURNAL	1 (bases 1 to 1837)		
REMARK	Strausberg, R.		
COMMENT	Direct Submission		
	Submitted (09-APR-2001) National Institutes of Health, Mammalian		
	Gene Collection (MGC), Cancer Genomics Office, National Cancer		
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		
	USA		
	NIH-MGC Project URL: http://mgc.nci.nih.gov		
	Contact: MGC help desk		
	Email: cgabs-remail.nih.gov		
	Tissue Procurement: Louis Staudt		
	cDNA Library Preparation: Rubin Laboratory		
	DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: National Institutes of Health Intramural		
	Sequencing Center (NISC),		
	Gaithersburg, Maryland;		
	Web site: http://www.nisc.nih.gov/		
	Contact: nisc.mgc@hgrl.nih.gov		
	Shenchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,		
	Benjamin, B., Blakesley, R.W., Bouffard, C.G., Brinkley, C., Brooks, S.,		
	Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Laquer, R.,		
	Lim, M., Maduro, O.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,		
	McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,		
	Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,		
	Zhang, L.-H. and Green, E.D.		
	Clone distribution: MGC clone distribution information can be found		
	through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov		
	Series: IRAL Plate: 17 Row: a Column: 1		
	This clone was selected for full length sequencing because it		
	passed the following selection criteria: matched mRNA q1: 9988409.		
FEATURES			
source	Location/Qualifiers		
	1..1837		
	/organism: "Homo sapiens"		
	/db_xref: "LocusID:57379"		
	/db_xref: "taxon:9606"		
	/clone: "MGC:12911 IMAGE:4054915"		
	/tissue_type: "Primary B-Cells from Tonsils"		
	/clone_lib: "NIH_MGC_48"		
	/lab_host: "DH10B-R"		
	/note: "Vector: pOT87"		
	86..682		
	/codon_start: 1		
	/product: "activation-induced cytidine deaminase"		
	/protein_id: "AAH06296.1"		
	/db_xref: "GI:13623401"		
	/translation: "MDSLIMNRKKFLYQFNVRWAKRRRETYICVYVVKRDSATSFSL		
	DFGYLRNKGCHVLELLFLRYISDWLDPCRCYRVTFWTSWSPCYDCARIVADPLKCNP		
	NLSLRFTFARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYTCWNTVENHRTFFK		
	ANEGLHNSVRLSRQLRLILPLVEVDDLRDAFRTLGL"		
BASE COUNT	530 a 387 c 421 g 499 t		
ORIGIN			

Query Match	64.4%	Score 1815;	DB 9;	Length 1837;
Best Local Similarity	100.0%;	Prod. No. 0;		
Matches 1815;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	4	GAACCATCATTAATTTGAAGTGTAGATTTTCTGGCCTGAGACTTTCGACGGAGGCAAGAAGA	63	
DB	10	GAACCATCATTAATTTGAAGTGTAGATTTTCTGGCCTGAGACTTTCGACGGAGGCAAGAAGA	69	
QY	64	CACCTCTGGACACCACTATGAGACGCCCTCTGTATGAACCGGAGGAAGTTTCTTTACCAATT	123	
DB	70	CACCTCTGGACACCACTATGAGACGCCCTCTGTATGAACCGGAGGAAGTTTCTTTACCAATT	129	
QY	124	CAAAAAATGTCGCTGGCTAAGGTGGCGCTGAGACCTACCTGTGCTACCTAGTGTGAAGAG	183	
DB	130	CAAAAAATGTCGCTGGCTAAGGTGGCGCTGAGACCTACCTGTGCTACCTAGTGTGAAGAG	189	
QY	184	GCCTGACAGTGTACATPCCCTTTTTCACCTGGACTTTTGGTTATCTTGCCTAATGAACACGGCTG	243	
DB	190	GCCTGACAGTGTACATPCCCTTTTTCACCTGGACTTTTGGTTATCTTGCCTAATGAACACGGCTG	249	
QY	244	CCAGCTGGAATTGCTCTTCTCCGCTACATCTCGGACTCGGACTGGACCTAGACCCCTGGCCCGCTG	303	
DB	250	CCAGCTGGAATTGCTCTTCTCCGCTACATCTCGGACTCGGACTGGACCTAGACCCCTGGCCCGCTG	309	
QY	304	CTACCGCGTCCACCTGGTTCACCTCCTCGAGCGCCCTGCTACGACTGTGCCCGACATGTGGC	363	
DB	310	CTACCGCGTCCACCTGGTTCACCTCCTCGAGCGCCCTGCTACGACTGTGCCCGACATGTGGC	369	
QY	364	CGACTTCTTCGGAGGGAACCCAACTCAGTCTGAGGATCTTACCCGGCGGCTCTACTT	423	
DB	370	CGACTTCTTCGGAGGGAACCCAACTCAGTCTGAGGATCTTACCCGGCGGCTCTACTT	429	
QY	424	CTGTGAGGACCGCAAGGCTGAGCCGAGGGCTGCGCGGCTGCACCGCGCGGGGTGCA	483	
DB	430	CTGTGAGGACCGCAAGGCTGAGCCGAGGGCTGCGCGGCTGCACCGCGCGGGGTGCA	489	
QY	484	AATAGCCATCATGACCTTCAAAGATTAATTTTACTGCTGGAACTACTTTGTAGAAAAACCA	543	
DB	490	AATAGCCATCATGACCTTCAAAGATTAATTTTACTGCTGGAACTACTTTGTAGAAAAACCA	549	
QY	544	TGAAGAACCTTTCAAAGCCCTGGAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACA	603	
DB	550	TGAAGAACCTTTCAAAGCCCTGGAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACA	609	
QY	604	GCCTTGGCGCATCCTTTTGGCCCTGTATGAGGTGTAGTACTTACGAGACGCAATTCGTAC	663	
DB	610	GCCTTGGCGCATCCTTTTGGCCCTGTATGAGGTGTAGTACTTACGAGACGCAATTCGTAC	669	
QY	664	TTTGGGACTTTGATAGCAACTTCCAGGAATGTACACACGATGAAATATCTCTGCTCAAG	723	
DB	670	TTTGGGACTTTGATAGCAACTTCCAGGAATGTACACACGATGAAATATCTCTGCTCAAG	729	
QY	724	ACAGTGGATAAAAACAGTCTCTCAAGTCTCTCTGTTTATTTCTTCAACTCTCACTTT	783	
DB	730	ACAGTGGATAAAAACAGTCTCTCAAGTCTCTCTGTTTATTTCTTCAACTCTCACTTT	789	
QY	784	CTTAGAGTTTACAGAAAAATATTTATATAGGACTCTTTAAAAAGATCTATGCTTGAAGA	843	
DB	790	CTTAGAGTTTACAGAAAAATATTTATATAGGACTCTTTAAAAAGATCTATGCTTGAAGA	849	
QY	844	ATAGAGAAGGAACACAGGCTTGGCCAGGGACGTGCTCAATTTGGTGCAGTTTTGAATGCA	903	
DB	850	ATAGAGAAGGAACACAGGCTTGGCCAGGGACGTGCTCAATTTGGTGCAGTTTTGAATGCA	909	
QY	904	ACATTGTCCCTTACTTGGGAATAACAGAACTGCAGGACCTGGGACATCTCTAAAGTGTCAA	963	
DB	910	ACATTGTCCCTTACTTGGGAATAACAGAACTGCAGGACCTGGGACATCTCTAAAGTGTCAA	969	
QY	964	CGTTTTTCTATGACTTTTAGTGTAGGTATGACAGCAGAGGATAGATCTCTTAAAAAGCATGGT	1023	
DB	970	CGTTTTTCTATGACTTTTAGTGTAGGTATGACAGCAGAGGATAGATCTCTTAAAAAGCATGGT	1029	

REFERENCE 2 (bases 1 to 596)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
Location/Qualifiers
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="2"
/cell_type="hybridoma P1-5"
/transgenic
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95416.1"
/db_xref="GI:22297246"
/translation="MDLLMNRKFLYOFKNRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELLFURYISDWLDPGRYRVTWFTSWSPCYDCARHVAFLRGNP
NLSLRIFARLYFCEDRKAEPGLRRLHRAVQIAIMTFKDYFCWNTFVENHIEHTFK
AWEGLHENSVRSLRLLPLLYEVDLDRDAFRTLGL"
BASE COUNT 128 a 163 c 155 g 150 t

Query Match 20.8%; Score 587; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.2e-294;
Matches 587; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 89 CTCCTTGATGAACCGGAGGAAGTTCTTTACCAATTCACAAATGTCGCGTGGCTAAGGT 148
Db 10 CTCCTTGATGAACCGGAGGAAGTTCTTTACCAATTCACAAATGTCGCGTGGCTAAGGT 69
QY 149 CGCGCTGAGACCTACCTGTGTACGTAGTGAAGAGCGGTGACAGTGTACATCCTTTTCA 208
Db 70 CGCGCTGAGACCTACCTGTGTACGTAGTGAAGAGCGGTGACAGTGTACATCCTTTTCA 129
QY 209 CTGGACTTTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGGAATTCCTTCCTCGGC 268
Db 130 CTGGACTTTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGGAATTCCTTCCTCGGC 189
QY 269 TACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCGTGTTCACCTCC 328
Db 190 TACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCGTGTTCACCTCC 249
QY 329 TGGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTCTGCGAGGGAACCCCAAC 388
Db 250 TGGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTCTGCGAGGGAACCCCAAC 309
QY 389 CTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGCTGAGCCC 448
Db 310 CTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGCTGAGCCC 369
QY 449 GAGGGCTGCGCGCGCTGACCGCGCGGGGTGCAAAATAGCCATCATGACCTTCAAGAT 508
Db 370 GAGGGCTGCGCGCGCTGACCGCGCGGGGTGCAAAATAGCCATCATGACCTTCAAGAT 429
QY 509 TATTTTTACTGCTGGAATACTTTTGTAGAAACCATGAAGAACCTTTCAAAGCCTGGAA 568
Db 430 TATTTTTACTGCTGGAATACTTTTGTAGAAACCATGAAGAACCTTTCAAAGCCTGGAA 489
QY 569 GGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCCTTTTGCCCGCTG 628
Db 490 GGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCCTTTTGCCCGCTG 549
QY 629 TATGAGTTGATGACTTACGAGACGCAATTCGTACTTTTGGGACTTTG 675
Db 550 TATGAGTTGATGACTTACGAGACGCAATTCGTACTTTTGGGACTTTG 596

RESULT 6
AF529823
LOCUS Homo sapiens clone Ramos 9 AID (AID) mRNA linear PRI 19-AUG-2002
DEFINITION AF529823
ACCESSION AF529823
VERSION AF529823.1 GI:22297233
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
Location/Qualifiers
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 9"
1..>596
/gene="AID"
1..>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95410.1"
/db_xref="GI:22297234"
/translation="MDSLMMNRKFLYOFKNRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELLFURYISDWLDPGRYRVTWFTSWSPCYDCARHVAFLRGNP
NLSLRIFARLYFCEDRKAEPGLRRLHRAVQIAIMTFKDYFCWNTFVENHIEHTFK
AWEGLHENSVRSLRLLPLLYEVDLDRDAFRTLGL"
BASE COUNT 128 a 163 c 155 g 150 t
ORIGIN
Query Match 20.7%; Score 584; DB 9; Length 596;
Best Local Similarity 100.0%; Pred. No. 4.5e-293;
Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 92 TTGATGAACCGGAGGAAGTTCTTTTACCAATTCACAAATGTCGCGTGGCTAAGGTGCG 151
Db 13 TTGATGAACCGGAGGAAGTTCTTTTACCAATTCACAAATGTCGCGTGGCTAAGGTGCG 72
QY 152 CGTGAGACCTACCTGTGTACGTAGTGAAGAGCGGTGACAGTGTACATCCTTTTCACTG 211
Db 73 CGTGAGACCTACCTGTGTACGTAGTGAAGAGCGGTGACAGTGTACATCCTTTTCACTG 132
QY 212 GACTTTGTTATCTTCGCAATGAAGACGGCTGCCACGTGGGAATTCCTTCCTCGCTAC 271
Db 133 GACTTTGTTATCTTCGCAATGAAGACGGCTGCCACGTGGGAATTCCTTCCTCGCTAC 192
QY 272 ATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCCCGTCACCTGGTTCACCTTCG 331
Db 193 ATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCCCGTCACCTGGTTCACCTTCG 252
QY 332 AGCCCTCTACGACTGTGCCGACATGTGCCGACTTTCTCGGAGGGAACCCCAACCTC 391
Db 253 AGCCCTCTACGACTGTGCCGACATGTGCCGACTTTCTCGGAGGGAACCCCAACCTC 312
QY 392 AGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGTTCAGTTCGAG 451
Db 313 AGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGTTCAGTTCGAG 472
QY 452 GGGCTGCGCGCGCTGCGACCCCGCGGGGTGCAAAATAGCCATCATGACCTTCAAGATTTAT 511

```
Db 373 GGGCTGCGCGGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACCTTCAAAGATTAT 432
QY 512 TTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAGAACTTTCAAAGCCTGGGAAGG 571
Db 433 TTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAGAACTTTCAAAGCCTGGGAAGG 492
QY 572 CTGCATGAAAATTCAGTTTCGCTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGCCCTGTAT 631
Db 493 CTGCATGAAAATTCAGTTTCGCTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGCCCTGTAT 552
QY 632 GAGTTGTAGTACTTACGAGAGCGCATTTTCGTACTTTGGGACTTTG 675
Db 553 GAGTTGTAGTACTTACGAGAGCGCATTTTCGTACTTTGGGACTTTG 596

RESULT 7
AF529846
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Chinese hamster.
Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus;
1 (bases 1 to 597)
Martin.A. and Scharff,M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Source
1. .597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="6"
/cell_line="CHO"
/transgenic
1. .597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .597
/gene="AID"
1. .597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95433.1"
/db_xref="GI:22297280"
/translation="MDSLMMNRKFLYQFNVRWAKRRRTYLCYVWKRDSATSFSL
DFGLRNKGCHVELLELYRISDWLDPRCYRVTWFTSWSPCYDCARHVDLFRGNP
NLSLRIFTARLYFCEDRAEPGLRLHRAGVQIAIMTFKDYFYCWNTFVENHRTFK
AWGLHENSURLSRQLRLPLLYEVDLDRAFRTWGR"
BASE COUNT 128 a 165 c 157 g 147 t
ORIGIN

Query Match 20.3%; Score 572; DB 12; Length 597;
Best Local Similarity 100.0%; Pred. No. 8.2e-287;
Matches 572; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 GATGAACCGGAGGAGTTTCTTTTACCAATTCAAAATGTCGCTGGGCTAAGGGTCGGCG 153
Db 15 GATGAACCGGAGGAGTTTCTTTTACCAATTCAAAATGTCGCTGGGCTAAGGGTCGGCG 74
QY 154 TGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACATCCTTTTCACTGGA 213
```

```
Db 75 TGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACATCCTTTTCACTGGA 134
QY 214 CTTTGGTTATCTTCGCRAATAGAACGGCTGCCAGTGGAAATTCCTTCCTCCGCTACAT 273
Db 135 CTTTGGTTATCTTCGCRAATAGAACGGCTGCCAGTGGAAATTCCTTCCTCCGCTACAT 194
QY 274 CTCGAGCTGGGACCTAGACCCCTGGCCGCTGTACCGGCTCACCTGGTTTCACCTCCTCGAG 333
Db 195 CTCGAGCTGGGACCTAGACCCCTGGCCGCTGTACCGGCTCACCTGGTTTCACCTCCTCGAG 254
QY 334 CCCCTGTACGACTGTGCCGACATGTGGCGGACTTTCTGCGAGGGAACCCCAACCTCAG 393
Db 255 CCCCTGTACGACTGTGCCGACATGTGGCGGACTTTCTGCGAGGGAACCCCAACCTCAG 314
QY 394 TCTGAGGATCTTCACCGCGGCCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGAGGG 453
Db 315 TCTGAGGATCTTCACCGCGGCCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGAGGG 374
QY 454 GCTGCGGGGCTGCACCGCGCGGGGTGCAATAGCCATCATGACCTTTCAAAGATTATTT 513
Db 375 GCTGCGGGGCTGCACCGCGCGGGGTGCAATAGCCATCATGACCTTTCAAAGATTATTT 434
QY 514 TTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTTTCAAAGCCTGGGAAGGCT 573
Db 435 TTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTTTCAAAGCCTGGGAAGGCT 494
QY 574 GCATGAAAATTCAGTTTCGCTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGCCCTGTATGA 633
Db 495 GCATGAAAATTCAGTTTCGCTCTCTCCAGACAGCTTCGGCGCATCCTTTTGGCCCTGTATGA 554
QY 634 GGTGTGATGACTTACGAGAGCGCATTTTCGTACTT 665
Db 555 GGTGTGATGACTTACGAGAGCGCATTTTCGTACTT 586

RESULT 8
AF529842
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Chinese hamster.
Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus;
1 (bases 1 to 597)
Martin.A. and Scharff,M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. .597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="2"
/cell_line="CHO"
/transgenic
1. .597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .597
/gene="AID"
1. .597
/gene="AID"

Query Match 20.3%; Score 572; DB 12; Length 597;
Best Local Similarity 100.0%; Pred. No. 8.2e-287;
Matches 572; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 GATGAACCGGAGGAGTTTCTTTTACCAATTCAAAATGTCGCTGGGCTAAGGGTCGGCG 153
Db 15 GATGAACCGGAGGAGTTTCTTTTACCAATTCAAAATGTCGCTGGGCTAAGGGTCGGCG 74
QY 154 TGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACATCCTTTTCACTGGA 213
```


/codon_start=1
/product="AID"
/protein_id="AA095429.1"
/db_xref="GI:22297272"
/translation="MDSLMMRRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFURYISDMDLDPGRCYRVTWFTSWSPCYDCARHVADEFURGNP
NLSLRIFTARLYFCEDRKAEPGLRRLHRAGVQIALMTFKDYFCWNTFVENHETFK
AWEGHENSVRSLRQLRLLPLVEVDLDRDAFRTWGR"
BASE COUNT 127 a 164 c 157 g 149 t
ORIGIN

Query Match 20.2%; Score 569; DB 12; Length 597;
Best Local Similarity 100.0%; Pred. No. 3e-285;
Matches 569; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 80 ATGGACAGCCCTTCATGAACGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGG 139
Db 1 ATGGACAGCCCTTCATGAACGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGG 60
Qy 140 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 199
Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120
Qy 200 TCCTTTTTCACGTGACCTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGGAAATGTCTC 259
Db 121 TCCTTTTTCACGTGACCTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGGAAATGTCTC 180
Qy 260 TTCTCTCCGCTACATCTCGGACCTGCGACCTAGACCTGCGCGCTGCTACCGGCTACCTGG 319
Db 181 TTCTCTCCGCTACATCTCGGACCTGCGACCTAGACCTGCGCGCTGCTACCGGCTACCTGG 240
Qy 320 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTTGGAGGG 379
Db 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTTGGAGGG 300
Qy 380 RACCCCAACCTCAGTCTGAGNATCTTACCGGGCGCCCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 RACCCCAACCTCAGTCTGAGNATCTTACCGGGCGCCCTCTACTTCTGTGAGGACCGCAAG 360
Qy 440 GCTGAGCCGAGGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 499
Db 361 GCTGAGCCGAGGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 420
Qy 500 TTCAAAGATTATTTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTTCAA 559
Db 421 TTCAAAGATTATTTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTTCAA 480
Qy 560 GCCTGGGAAGGCTGCATGAATAATTCAGTTCGTCTCTCCAGACACTTTCGGCGCATCCTT 619
Db 481 GCCTGGGAAGGCTGCATGAATAATTCAGTTCGTCTCTCCAGACACTTTCGGCGCATCCTT 540
Qy 620 TTGCCCCCTGTATGAGTTGTGATCACTTACG 648
Db 541 TTGCCCCCTGTATGAGTTGTGATCACTTACG 569

RESULT 9
AF529847
LOCUS AF529847 597 bp mRNA linear SYN 19-AUG-2002
DEFINITION Cricetus griseus clone 7 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
ACCESSION AF529847
VERSION AF529847.1 GI:22297281
KEYWORDS Chinese hamster.
SOURCE Cricetus griseus
ORGANISM Cricetus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells

Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Location/Qualifiers
1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="7"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AA095434.1"
/db_xref="GI:22297282"
/translation="MDSLMMRRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFURYISDMDLDPGRCYRVTWFTSWSPCYDCARHVADEFURGNP
NLSLRIFTARLYFCEDRKAEPGLRRLHRAGVQIALMTFKDYFCWNTFVENHETFK
AWEGHENSVRSLRQLRLLPLVEVDLDRDAFRTWGR"
BASE COUNT 129 a 164 c 156 g 148 t
ORIGIN
Query Match 19.4%; Score 548; DB 12; Length 597;
Best Local Similarity 100.0%; Pred. No. 2.7e-274;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 80 ATGGACAGCCCTTCGATGAACGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGGCTGG 139
Db 1 ATGGACAGCCCTTCGATGAACGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGGCTGG 60
Qy 140 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 199
Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120
Qy 200 TCCTTTTTCACGTGACCTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGGAAATGTCTC 259
Db 121 TCCTTTTTCACGTGACCTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGGAAATGTCTC 180
Qy 260 TTCTCTCCGCTACATCTCGGACCTGCGACCTAGACCTGCGCGCTGCTACCGCTACCTGG 319
Db 181 TTCTCTCCGCTACATCTCGGACCTGCGACCTAGACCTGCGCGCTGCTACCGCTACCTGG 240
Qy 320 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTTGGAGGG 379
Db 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTTGGAGGG 300
Qy 380 AACCCCAACCTCAGTCTGAGNATCTTACCGCGCGGGTGCAATAGCCATCATGACC 439
Db 301 AACCCCAACCTCAGTCTGAGNATCTTACCGCGCGGGTGCAATAGCCATCATGACC 360
Qy 440 GCTGAGCCGAGGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 499
Db 361 GCTGAGCCGAGGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 420
Qy 500 TTCAAAGATTATTTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTTCAA 559
Db 421 TTCAAAGATTATTTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTTCAA 480
Qy 560 GCCTGGGAAGGCTGCATGAATAATTCAGTTCGTCTCTCCAGACACTTTCGGCGCATCCTT 619
Db 481 GCCTGGGAAGGCTGCATGAATAATTCAGTTCGTCTCTCCAGACACTTTCGGCGCATCCTT 540
Qy 620 TTGCCCCCT 627
|||||||

gene /db_xref="taxon:9606"
1. .>596
/clone="Ramos 5"
CDS /gene="AID"
1. .>596
/gene="AID"
/note="integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95406.1"
/db_xref="GI:22297226"
/translation="MDSLLNRRKFLYOFKNVRWAKRRRETYLCYVVKRRDSATSFSL
DFGLRNKNCHVELLEFLRYISDWLDLPGRYRVTFWSWPCYDCARHVDLFRGNP
NLSIRIFTARLYCFDKHAPKPEGLRLHRAGVOIAIMTKDFYCNWTFVENHERTEFK
AWEGLHNSVRLSRQLRILLPLYEVDLDRDAFTLGL"
BASE COUNT 128 a 164 c 154 g 150 t
ORIGIN

Query Match 19.3%; Score 545; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 80 ATGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
Qy 140 GCTAAGGTCGGCTGAGACCTACTGTCTACGTAGTGAAGAGCGCTGACAGTCTACA 199
Db 61 GCTAAGGTCGGCTGAGACCTACTGTCTACGTAGTGAAGAGCGCTGACAGTCTACA 120
Qy 200 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGAAATGCTC 259
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGAAATGCTC 180
Qy 260 TTCCTCCGTACATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGCGTCACCTGG 319
Db 181 TTCCTCCGTACATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGCGTCACCTGG 240
Qy 320 TTCACCTCCTGGAGCCCTGCTACGACTGTGCCCGAGATGTGGCGGACTTTCTCGGAGGG 379
Db 241 TTCACCTCCTGGAGCCCTGCTACGACTGTGCCCGAGATGTGGCGGACTTTCTCGGAGGG 300
Qy 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Qy 440 GCTGAGCCCGAGGGCTGCGCGGCTGCACCGCGCGGCTGCAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGCTGCGCGGCTGCACCGCGCGGCTGCAATAGCCATCATGACC 420
Qy 500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAAGAACTTTCAAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAAGAACTTTCAAA 480
Qy 560 GCCTGGGAAGGGCTGCATGAAAATTCAGTTCGTCTCCAGACAGCTTCGGCGCATCCTT 619
Db 481 GCCTGGGAAGGGCTGCATGAAAATTCAGTTCGTCTCCAGACAGCTTCGGCGCATCCTT 540
Qy 620 TTGCCCTGTATGAGTTGATGACTTACGAGACCGCATTTCTACTTTGGGACTTTG 675
Db 541 TTGCCCTGTATGAGTTGATGACTTACGAGACCGCATTTCTACTTTGGGACTTTG 596

RESULT 14
AF529820 596 bp mRNA linear PRI 19-AUG-2002
LOCUS Homo sapiens clone Ramos 6 AID (AID) mRNA, partial cds.
DEFINITION AF529820
ACCESSION AF529820
VERSION AF529820.1 GI:22297227
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL .Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1. .596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 6"
1. .>596
/gene="AID"
/note="integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95407.1"
/db_xref="GI:22297228"
/translation="MDSLLNRRKFLYOFKNVRWAKRRRETYLCYVVKRRDSATSFSL
DFGLRNKNCHVELLEFLRYISDWLDLPGRYRVTFWSWPCYDCARHVDLFRGNP
NLSIRIFTARLYCFDKHAPKPEGLRLHRAGVOIAIMTKDFYCNWTFVENHERTEFK
AWEGLHNSVRLSRQLRILLPLYEVDLDRDAFTLGL"
BASE COUNT 126 a 164 c 155 g 151 t
ORIGIN
Query Match 19.3%; Score 545; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 80 ATGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
Qy 140 GCTAAGGTCGGCTGAGACCTACTGTCTACGTAGTGAAGAGCGCTGACAGTCTACA 199
Db 61 GCTAAGGTCGGCTGAGACCTACTGTCTACGTAGTGAAGAGCGCTGACAGTCTACA 120
Qy 200 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGAAATGCTC 259
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGAAATGCTC 180
Qy 260 TTCCTCCGTACATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGCGTCACCTGG 319
Db 181 TTCCTCCGTACATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGCGTCACCTGG 240
Qy 320 TTCACCTCCTGGAGCCCTGCTACGACTGTGCCCGAGATGTGGCGGACTTTCTCGGAGGG 379
Db 241 TTCACCTCCTGGAGCCCTGCTACGACTGTGCCCGAGATGTGGCGGACTTTCTCGGAGGG 300
Qy 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Qy 440 GCTGAGCCCGAGGGCTGCGCGGCTGCACCGCGCGGCTGCAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGCTGCGCGGCTGCACCGCGCGGCTGCAATAGCCATCATGACC 420
Qy 500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAAGAACTTTCAAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAAGAACTTTCAAA 480
Qy 560 GCCTGGGAAGGGCTGCATGAAAATTCAGTTCGTCTCCAGACAGCTTCGGCGCATCCTT 619
Db 481 GCCTGGGAAGGGCTGCATGAAAATTCAGTTCGTCTCCAGACAGCTTCGGCGCATCCTT 540
Qy 620 TTGCCCTGTATGAGTTGATGACTTACGAGACCGCATTTCTACTTTGGGACTTTG 675

```
|||||
541 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTTGGGACTTTG 596

RESULT 15
AF529821
LOCUS Homo sapiens clone Ramos 7 AID (AID) mRNA, partial cds. PRI 19-AUG-2002
DEFINITION
ACCESSION AF529821
VERSION AF529821.1 GI:22297229
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 596)
Martin, A. and Scharff, M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source
1. 596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 7"
1. >596
/gene="AID"
1. >596
/gene="AID"
/feature="AID"
/feature="integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95408.1"
/db_xref="GI:22297230"
/translation="MDSLMMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDRATSFSL
DFGLRNKNGCHVELLFLRYISDNDLPDGRGYRTWFTSNPCYDCARHVDPLRGNP
NLSURIFATRIYFCEDRKAPEGLRLHRAVGQIATMTFKDYFCWNTFVENHIERFK
AWEGHENSVRSLRLLPLLYEVDLDRAFTLGL"
BASE COUNT 127 a 165 c 155 g 149 t
ORIGIN

Query Match 19.3%; Score 545; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 80 ATGGACAGCCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGGACAGCCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGCTGG 60
Qy 140 GCTAAGGTCGGCTGACACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTCTACA 199
Db 61 GCTAAGGTCGGCTGACACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTCTACA 120
Qy 200 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCACGTGAATTGCTC 259
Db 121 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCACGTGAATTGCTC 180
Qy 260 TTCCTCCGCTACATCTCGACTGGGACCTAGACCCCTGGCCGCTGTACCGGCTCACCTGG 319
Db 181 TTCCTCCGCTACATCTCGACTGGGACCCAGACCCCTGGCCGCTGTACCGGCTCACCTGG 240
Qy 320 TTCACCTCTGAGAGCCCTGCTACGACTGTGCCGACATGTGGCGGCACTTTCTGCGAGGG 379
Db 241 TTCACCTCTGAGAGCCCTGCTACGACTGTGCCGACATGTGGCGGCACTTTCTGCGAGGG 300
Qy 380 AACCCCAACCTCAGTCTCAGATCTTACCGCGCCCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTCAGATCTTACCGCGCCCTCTACTTCTGTGAGGACCGCAAG 360
```

```
|||||
440 GCTGAGCCCGAGGGGCTGGCGCTGCACCGCCCGGGTGCAGAAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGGCTGGCGCTGCACCGCCCGGGTGCAGAAATAGCCATCATGACC 420
Qy 500 TTCAAAGATTATTTTACTGCTGGAATACATTTTCTAGAAAAACCATGAAGAACTTTCAAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATACATTTTCTAGAAAAACCATGAAGAACTTTCAAA 480
Qy 560 GCCTGGGAAGGCGTGCATGAAAAATTCAGTTCTCTCCAGACAGCTTTCGCCGATCCTT 619
Db 481 GCCTGGGAAGGCGTGCATGAAAAATTCAGTTCTCTCCAGACAGCTTTCGCCGATCCTT 540
Qy 620 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 675
Db 541 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 596

RESULT 16
AF529822
LOCUS Homo sapiens clone Ramos 8 AID (AID) mRNA, partial cds. PRI 19-AUG-2002
DEFINITION
ACCESSION AF529822
VERSION AF529822.1 GI:22297231
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 596)
Martin, A. and Scharff, M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source
1. 596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 8"
1. >596
/gene="AID"
1. >596
/gene="AID"
/feature="AID"
/feature="integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95409.1"
/db_xref="GI:22297232"
/translation="MDSLMMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDRATSFSL
DFGLRNKNGCHVELLFLRYISDNDLPDGRGYRTWFTSNPCYDCARHVDPLRGNP
NLSURIFATRIYFCEDRKAPEGLRLHRAVGQIATMTFKDYFCWNTFVENHIERFK
AWEGHENSVRSLRLLPLLYEVDLDRAFTLGL"
BASE COUNT 126 a 165 c 155 g 150 t
ORIGIN

Query Match 19.3%; Score 545; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 80 ATGGACAGCCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGGACAGCCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGCTGG 60
Qy 140 GCTAAGGTCGGCTGACACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTCTACA 199
Db 61 GCTAAGGTCGGCTGACACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTCTACA 120
Qy 200 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCACGTGAATTGCTC 259
```

```

121 TCCCTTTTCTGCTGAGCTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGCTGGAATTTGCTC 180
260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTTGGCCGCTGCTACCGCGTCACTCTGG 319
181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTTGGCCGCTGCTACCGCGTCACTCTGG 240
320 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATCTGGCCGACTTTCTTGGCAGGG 379
241 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATCTGGCCGACTTTCTTGGCAGGG 300
380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
440 GCTGAGCCCGAGGGGCTGCGGGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC 499
361 GCTGAGCCCGAGGGGCTGCGGGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC 420
500 TTCAAAGATTTATTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAAGAACTTTCAAA 559
421 TTCAAAGATTTATTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAAGAACTTTCAAA 480
560 GCTGGAAGGGCTGCATGAATTCAGTTCGCTCTCCAGACAGCTTTCGGCGCATCTT 619
481 GCTGGAAGGGCTGCATGAATTCAGTTCGCTCTCCAGACAGCTTTCGGCGCATCTT 540
620 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTTGCTACTTTGGGACTTTG 675
541 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTTGCTACTTTGGGACTTTG 596

RESULT 17
AF529824
LOCUS Homo sapiens clone Ramos 10 AID (AID) mRNA, partial cds. PRI 19-AUG-2002
DEFINITION AF529824
ACCESSION AF529824
VERSION AF529824.1 GI:22297235
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
SOURCE 1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 10"
1..>596
/genes="AID"
1..>596
/genes="AID"
/notes="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95411.1"
/db_xref="GI:22297236"
/translocation="MDSLLNMRKKFLYQKFNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGWVLELLFLRYISDWLDPGRCYRVTFWTSFPCYCARHVADEFLRGNP
NLSLRTFARLYFCEDRKAEPGLRHLRHAGVQIAIMTEKDFYCWNTVEVNHRTFK
AWEGLHNSVRLSRQLRLLPLLYEVDLLRDAFRTGL"
BASE COUNT 127 a 163 c 156 g 150 t
ORIGIN
```

```

Query Match 19.3%; Score 545; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGGCTCTTGATGAACCGGAGGAAGTTCTTTTACCAATTCAAAATGTCCGCTGG 139
Db 1 ATGGACAGGCTCTTGATGAACCGGAGGAAGTTCTTTTACCAATTCAAAATGTCCGCTGG 60
QY 140 GCTAAGGGTGCGGTGGAGACCTACCTGTGCTAGTAGTGAAGAGGGCTCACAGTGTCTACA 199
Db 61 GCTAAGGGTGCGGTGGAGACCTACCTGTGCTAGTAGTGAAGAGGGCTCACAGTGTCTACA 120
QY 200 TCTCTTTTCACTGGACTTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGCTGGAATTTGCTC 259
Db 121 TCTCTTTTCACTGGACTTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGCTGGAATTTGCTC 180
QY 260 TTCTCTCCGCTACATCTCGGACTGGGACCTAGACCTTGGCCGCTGCTACCGCGTCACTCTGG 319
Db 181 TTCTCTCCGCTACATCTCGGACTGGGACCTAGACCTTGGCCGCTGCTACCGCGTCACTCTGG 240
QY 320 TTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTTCTGCGAGGG 379
Db 241 TTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTTCTGCGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCCGAGGGGCTGCGGGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGGCTGCGGGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC 420
QY 500 TTCAAGATTTATTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAAGAACTTTCAAA 559
Db 421 TTCAAGATTTATTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAAGAACTTTCAAA 480
QY 560 GCTTGGGAAGGGCTGCATGAATTCAGTTCGCTCTCCAGACAGCTTTCGGCGCATCTT 619
Db 481 GCTTGGGAAGGGCTGCATGAATTCAGTTCGCTCTCCAGACAGCTTTCGGCGCATCTT 540
QY 620 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTTGCTACTTTGGGACTTTG 675
Db 541 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTTGCTACTTTGGGACTTTG 596

RESULT 18
AF529825
LOCUS Homo sapiens clone Ramos 11 AID (AID) mRNA, partial cds. PRI 19-AUG-2002
DEFINITION AF529825
ACCESSION AF529825
VERSION AF529825.1 GI:22297237
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
SOURCE 1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 11"
1..>596
gene
```

```
CDS
/feature="AID"
1..>596
/feature="AID"
/feature="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95412.1"
/db_xref="GI:22297238"
/translation="MDSLIMNRKFLYQFNVRWAKGRREAYLCYVVKRDSATSFSL
DFGYLRNKGCHVELLFYRISDWLDLPGRCYRVTFWTSWSPCYDCARHVAFLRGNP
NLSLRFTARLYFCEDRKAEPGLRHRAGVQIAIMTFKDYFCWNTFVENHIERTEK
AWEGLHENSVRSLRQLRILLPLYEVDLDRDAFRTLGL"
BASE COUNT 126 a 164 c 156 g 150 t
ORIGIN
Query Match 19.3%; Score 545; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 80 ATGCACAGCCTCTTGATGAACCGAGGAAAGTTCTTTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGCACAGCCTCTTGATGAACCGAGGAAAGTTCTTTTACCAATTCAAAAATGTCGCTGG 60
Qy 140 GCTAAGGGTCGGCTGAGACCTACCTGTGCTAGTGAAGAGGGCTGACAGTGTCTACA 199
Db 61 GCTAAGGGTCGGCTGAGACCTACCTGTGCTAGTGAAGAGGGCTGACAGTGTCTACA 120
Qy 200 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGAAATGGCTC 259
Db 121 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGAAATGGCTC 180
Qy 260 TTCTCTCGCTACATCTCGGACTGGGACTGGGACTGGGACTGGGACTGGGACTGGGACTGG 240
Db 181 TTCTCTCGCTACATCTCGGACTGGGACTGGGACTGGGACTGGGACTGGGACTGGGACTGG 379
Qy 320 TTCAACCTCTCGGAGCCCTGCTACGACTGTGCGCGACATGTGCGCGACTTTCTGCGAGGG 300
Db 241 TTCAACCTCTCGGAGCCCTGCTACGACTGTGCGCGACATGTGCGCGACTTTCTGCGAGGG 439
Qy 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGAGGACCGCAAG 499
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGAGGACCGCAAG 420
Qy 440 GCTGAGCCCGAGGGCTGCGGCGCTGACCGCGCGCGCTGACCGCGCGCGCTGACCGCGCG 499
Db 361 GCTGAGCCCGAGGGCTGCGGCGCTGACCGCGCGCGCTGACCGCGCGCGCTGACCGCGCG 420
Qy 500 TTCAAAAGATTATTTTACTGCTGGAATACCTTCTAGAAAACCATGAAGACATTTTCAAA 559
Db 421 TTCAAAAGATTATTTTACTGCTGGAATACCTTCTAGAAAACCATGAAGACATTTTCAAA 480
Qy 560 GCCTGGGAAGGGCTGCAATGAAATTCAGTTTCTCTCCAGACAGCTTTGGCGCATCCTT 619
Db 481 GCCTGGGAAGGGCTGCAATGAAATTCAGTTTCTCTCCAGACAGCTTTGGCGCATCCTT 540
Qy 620 TTGCCCCCTGTATGAGTTGATGACTTACGAGACGCATTTTCTACTTTGGGACTTTG 675
Db 541 TTGCCCCCTGTATGAGTTGATGACTTACGAGACGCATTTTCTACTTTGGGACTTTG 596
RESULT 19
AF529826 596 bp mRNA linear PRI 19-AUG-2002
LOCUS Homo sapiens clone Ramos 12 AID (AID) mRNA, partial cds.
DEFINITION AF529826
ACCESSION AF529826
VERSION AF529826.1 GI:22297239
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
```

```
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 12"
1..>596
/feature="AID"
1..>596
/feature="AID"
/feature="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95413.1"
/db_xref="GI:22297240"
/translation="MDSLIMNRKFLYQFNVRWAKGRRETYLCYVVKRDSATSFSL
DFGYLRNKGCHVELLFYRISDWLDLPGRCYRVTFWTSWSPCYDCARHVAFLRGNP
NLSLRFTARLYFCEDRKAEPGLRHRAGVQIAIMTFKDYFCWNTFVENHIERTEK
AWEGLHENSVRSLRQLRILLPLYEVDLDRDAFRTLGL"
BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN
Query Match 19.3%; Score 545; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 80 ATGCACAGCCTCTTGATGAACCGAGGAAAGTTCTTTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGCACAGCCTCTTGATGAACCGAGGAAAGTTCTTTTACCAATTCAAAAATGTCGCTGG 60
Qy 140 GCTAAGGGTCGGCTGAGACCTACCTGTGCTAGTGAAGAGGGCTGACAGTGTCTACA 199
Db 61 GCTAAGGGTCGGCTGAGACCTACCTGTGCTAGTGAAGAGGGCTGACAGTGTCTACA 120
Qy 200 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGAAATGGCTC 259
Db 121 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGAAATGGCTC 180
Qy 260 TTCTCTCGCTACATCTCGGACTGGGACTGGGACTGGGACTGGGACTGGGACTGGGACTGG 319
Db 181 TTCTCTCGCTACATCTCGGACTGGGACTGGGACTGGGACTGGGACTGGGACTGGGACTGG 240
Qy 320 TTCAACCTCTCGGAGCCCTGCTACGACTGTGCGCGACATGTGCGCGACTTTCTGCGAGGG 379
Db 241 TTCAACCTCTCGGAGCCCTGCTACGACTGTGCGCGACATGTGCGCGACTTTCTGCGAGGG 300
Qy 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGAGGACCGCAAG 419
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGAGGACCGCAAG 360
Qy 440 GCTGAGCCCGAGGGCTGCGGCGCTGACCGCGCGCGCTGACCGCGCGCGCTGACCGCGCG 499
Db 361 GCTGAGCCCGAGGGCTGCGGCGCTGACCGCGCGCGCTGACCGCGCGCGCTGACCGCGCG 420
Qy 500 TTCAAAAGATTATTTTACTGCTGGAATACCTTCTAGAAAACCATGAAGACATTTTCAAA 559
Db 421 TTCAAAAGATTATTTTACTGCTGGAATACCTTCTAGAAAACCATGAAGACATTTTCAAA 480
Qy 560 GCCTGGGAAGGGCTGCAATGAAATTCAGTTTCTCTCCAGACAGCTTTGGCGCATCCTT 619
Db 481 GCCTGGGAAGGGCTGCAATGAAATTCAGTTTCTCTCCAGACAGCTTTGGCGCATCCTT 540
Qy 620 TTGCCCCCTGTATGAGTTGATGACTTACGAGACGCATTTTCTACTTTGGGACTTTG 675
Db 541 TTGCCCCCTGTATGAGTTGATGACTTACGAGACGCATTTTCTACTTTGGGACTTTG 596
```

```
RESULT 20
AF529827
LOCUS Homo sapiens clone Ramos 13 AID (AID) mRNA, partial cds.
DEFINITION Homo sapiens clone Ramos 13 AID (AID) mRNA, partial cds.
ACCESSION AF529827
VERSION AF529827.1 GI:22297241
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Location/Qualifiers
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 13"
1..>596
/gene="AID"
1..>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95414.1"
/db_xref="GI:22297242"
/translation="MDSLMMRRKFLYQFKVNRWAKGRRETYLCVWVKRRSATSFSL
DFGYLRNKGCHVELLFYISDWLDPRCYRVWTFWSPCYDCARHVDLFRGNP
NLSRIFATRLYFCEDRKAEPGLRRLHAGVQIAIMTFKDYFCWNTFVENHRTFFK
AWEGLHENSRLSLRLLPLLYEVDLROAFRTLGL"
BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN
Query Match 19.3%; Score 545; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTACCAATTCAAAAATGTCGCTGG 60
QY 140 GCTAAGGTCGCGGTGAGACCTTACCTGTGCTACGTAGTGAAGCGGTGACAGTGTCTACA 199
Db 61 GCTAAGGTCGCGGTGAGACCTTACCTGTGCTACGTAGTGAAGCGGTGACAGTGTCTACA 120
QY 200 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGAATGTCTC 259
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGAATGTCTC 180
QY 260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGTGTACCGGCTCACTGG 319
Db 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGTGTACCGGCTCACTGG 240
QY 320 TTCACCTCCTGGAGCCCTGCTTACGACTGTGCGCCGACATGTGCGCGACTTTCTGCGAGGG 379
Db 241 TTCACCTCCTGGAGCCCTGCTTACGACTGTGCGCCGACATGTGCGCGACTTTCTGCGAGGG 300
QY 380 ACCCCAACTCAGTCTGAGGATCTTACCGCGCGCCCTCTACTTCTGTGAGAGACCGCAAG 439
Db 301 AACCCCAACTCAGTCTGAGGATCTTATCGCGCGCCCTCTACTTCTGTGAGGAGCGCAAG 360
QY 440 GCTGAGCCCGGAGGCGCTGCGCGGCTGCACCCGCGGGGTGCAATAGCCATCATGACC 499
|||||
```

```
Db 361 GCTGAGCCCGAGGGCTCGCGCGGTGCACCCGCGGGTGCAAAATAGCCATCATGACC 420
QY 500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA 559
|||||
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA 480
QY 560 GCCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCCPT 619
Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCCPT 540
QY 620 TTGCCCCCTGTATGAGGTTGATGACATACGAGACGATTCGTTACGTTGGGACTTTG 675
Db 541 TTGCCCCCTGTATGAGGTTGATGACATACGAGACGATTCGTTACGTTGGGACTTTG 596
RESULT 21
AF529830
LOCUS Mus musculus clone 3 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529830
VERSION AF529830.1 GI:22297247
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Location/Qualifiers
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="3"
/cell_type="hybridoma P1-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95417.1"
/db_xref="GI:22297248"
/translation="MDSLMMRRKFLYQFKVNRWAKGRRETYLCVWVKRRSATSFSL
DFGYLRNKGCHVELLFYISDWLDPRCYRVWTFWSPCYDCARHVDLFRGNP
NLSRIFATRLYFCEDRKAEPGLRRLHAGVQIAIMTFKDYFCWNTFVENHRTFFK
AWEGLHENSRLSLRLLPLLYEVDLROAFRTLGL"
BASE COUNT 126 a 164 c 155 g 151 t
ORIGIN
Query Match 19.3%; Score 545; DB 12; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTACCAATTCAAAAATGTCGCTGG 60
QY 140 GCTAAGGTCGCGGTGAGACCTTACCTGTGCTACGTAGTGAAGCGGTGACAGTGTCTACA 199
Db 61 GCTAAGGTCGCGGTGAGACCTTACCTGTGCTACGTAGTGAAGCGGTGACAGTGTCTACA 120
```



```
QY 200 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCGTC 259
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCGTC 180
QY 260 TTCCTCCGCTACATCTCGGACTGGGACCTACACCTGGCCGCTGCTACCGGTCACCTGG 319
Db 181 TTCCTCCGCTACATCTCGGACTGGGACCTACACCTGGCCGCTGCTACCGGTCACCTGG 240
QY 320 TTCACCTCTCGAGGCGCTGCTACGACGTGCGCGACATGCGGACATGCTTCTCGAGGG 379
Db 241 TTCACCTCTCGAGGCGCTGCTACGACGTGCGCGACATGCGGACATGCTTCTCGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCGCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCGCTCTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCGAGGGCTCGCGGCTGACCGCGCGGCGGTGCAAAATGACCATCATGACC 499
Db 361 GCTGAGCCGAGGGCTCGCGGCTGACCGCGCGGCGGTGCAAAATGACCATCATGACC 420
QY 500 TTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAAGAACTTTTCAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAAGAACTTTTCAA 480
QY 560 GCTGGGAGGGCTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGCGCATCCTT 619
Db 481 GCTGGGAGGGCTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGCGCATCCTT 540
QY 620 TTGCCCTGTATGAGTTGATGACTTAGGACGCAATTCGTTTGGGACTTTG 675
Db 541 TTGCCCTGTATGAGTTGATGACTTAGGACGCAATTCGTTTGGGACTTTG 596

RESULT 22
AF529831
LOCUS AF529831 Mus musculus clone 4 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529831
VERSION AF529831.1 GI:22297249
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Martin, A. and Scharff, M.D.
JOURNAL Somatic hypermutation of the AID transgene in B and non-B cells
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="4"
/cell_type="hybridoma P1-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="Aid"
/codon_start=1
/product="AID"
/protein_id="AAM95418.1"
/db_xref="GI:22297250"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
```

```
DFGLRNKNGCHVELLFRLYISDWLDPGRCYRVMTFTSMSPCYDCAHIVDFLRGNP
NLSRIETARLYFCEDRRKAEPLRLRHAGVQIAIMTFKDYFYCWMTFVENIERTEK
AWEGLHNSVRLSRQFRILLPLLYEVDLRDAPRTIGL"
BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN
Query Match 19.3%; Score 545; DB 12; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGACAGCCCTCTTGATGTAACCGGAGGAAGTTTCTTTACCAATTCACAAATTCGTCGGTGG 139
Db 1 ATGACAGCCCTCTTGATGTAACCGGAGGAAGTTTCTTTACCAATTCACAAATTCGTCGGTGG 60
QY 140 GCTAAGGCTCGGCTGAGACCTACCTGCTGCTACCTAGTGAAGAGCGGTGACAGTGCATACA 199
Db 61 GCTAAGGCTCGGCTGAGACCTACCTGCTGCTACCTAGTGAAGAGCGGTGACAGTGCATACA 120
QY 200 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCGTC 259
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCGTC 180
QY 260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCGGTCACCTGG 319
Db 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCGGTCACCTGG 240
QY 320 TTCACCTCTCGAGGCGCTGCTACGACGTGCGCGACATGCGGCGGCTTCTTCTGCGAGGG 379
Db 241 TTCACCTCTCGAGGCGCTGCTACGACGTGCGCGACATGCGGCGGCTTCTTCTGCGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCGCTCTACTTCTGTGAGACGCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCGCTCTACTTCTGTGAGACGCGCAAG 360
QY 440 GCTGAGCCGAGGGCTCGCGGCTGACCGCGCGGCGGTGCAAAATGACCATCATGACC 499
Db 361 GCTGAGCCGAGGGCTCGCGGCTGACCGCGCGGCGGTGCAAAATGACCATCATGACC 420
QY 500 TTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAAGAACTTTTCAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAAGAACTTTTCAA 480
QY 560 GCTGGGAGGGCTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGCGCATCCTT 619
Db 481 GCTGGGAGGGCTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGCGCATCCTT 540
QY 620 TTGCCCTGTATGAGTTGATGACTTAGGACGCAATTCGTTTGGGACTTTG 675
Db 541 TTGCCCTGTATGAGTTGATGACTTAGGACGCAATTCGTTTGGGACTTTG 596

RESULT 23
AF529833
LOCUS AF529833 Mus musculus clone 6 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529833
VERSION AF529833.1 GI:22297253
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Martin, A. and Scharff, M.D.
JOURNAL Somatic hypermutation of the AID transgene in B and non-B cells
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
```



```
source      1. .596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="6"
/cell_type="hybridoma P1-5"
/transgenic

source      1. .596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="8"
/cell_type="hybridoma P1-5"
/transgenic

gene        1. .>596
/genes="AID"

CDS         1. .>596
/genes="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95420.1"
/db_xref="GI:22297254"
/translation="MDSLNNRRKFLYQKNVRWAKGRRETYVCYVVKRDSATSLSL
DFGYLRNKGCHVELLFLRYISDWLDLPGRCYRVWTFSWSPCYDCARHVADFLRNP
NLSIRFTARLYFEDRKAEPLRRLHRAVQIAIMTKDYFCWNTFVENHERTFK
AWBGLHNSVRLSRQLRILLPLYEVDLRLDAFRTLGL"

BASE COUNT 127 a 163 c 156 g 150 t

Query Match      19.3%; Score 545; DB 12; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 80 ATGACAGCCTCTTGATGAACCGGAGGAGTCTTTTACCAATTCAAAAATGCCGTGG 139
Db 1 ATGACAGCCTCTTGATGAACCGGAGGAGTCTTTTACCAATTCAAAAATGCCGTGG 60

Qy 140 GCTAAGGGTCGGGTGAGACCTACCTGTGTAGTAGTGAAGAGGGGTGACAGTGCTACA 199
Db 61 GCTAAGGGTCGGGTGAGACCTACCTGTGTAGTAGTGAAGAGGGGTGACAGTGCTACA 120

Qy 200 TCCTTTTCACTGACATTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCGCTC 259
Db 121 TCCTTTTCACTGACATTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCGCTC 180

Qy 260 TTCCTCCGCTACATCTCGGACTCGGACCTAGACCTGGCCGCTGCTACCCGCTCACCTGG 319
Db 181 TTCCTCCGCTACATCTCGGACTCGGACCTAGACCTGGCCGCTGCTACCCGCTCACCTGG 240

Qy 320 TTCACCTCCGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTCCTGCGAGGG 379
Db 241 TTCACCTCCGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTCCTGCGAGGG 300

Qy 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTCTGTGAGGACCGCAAG 360

Qy 440 GCTGAGCCCGAGGGGTGCGCGGCTGCACCGCGCGGGTGCATAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGGTGCGCGGCTGCACCGCGCGGGTGCATAATAGCCATCATGACC 420

Qy 500 TTCAGAATATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAGAACTTTCAA 559
Db 421 TTCAGAATATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAGAACTTTCAA 480

Qy 560 GCCTGGGAGGGGTGCATGAATTCAGTTTCGTTCTCTCCAGACAGCTTCGGCGCATCCTT 619
Db 481 GCCTGGGAGGGGTGCATGAATTCAGTTTCGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540

Qy 620 TTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTTCTTTGGGACTTTG 675
Db 541 TTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTTCTTTGGGACTTTG 596

RESULT 24
AF529835 596 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 8 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION
```

```
complete cds.
ACCESSION AF529835
VERSION AF529835.1 GI:22297257
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1. .596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="8"
/cell_type="hybridoma P1-5"
/transgenic

source      1. .596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="8"
/cell_type="hybridoma P1-5"
/transgenic

gene        1. .>596
/genes="AID"

CDS         1. .>596
/genes="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95422.1"
/db_xref="GI:22297258"
/translation="MDSLNNRRKFLYQKNVRWAKGRRETYVCYVVKRDSATSLSL
DFGYLRNKGCHVELLFLRYISDWLDLPGRCYRVWTFSWSPCYDCARHVADFLRNP
NLSIRFTARLYFEDRKAEPLRRLHRAVQIAIMTKDYFCWNTFVENHERTFK
AWBGLHNSVRLSRQLRILLPLYEVDLRLDAFRTLGL"

BASE COUNT 127 a 163 c 156 g 150 t

Query Match      19.3%; Score 545; DB 12; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 80 ATGACAGCCTCTTGATGAACCGGAGGAGTCTTTTACCAATTCAAAAATGCCGTGG 139
Db 1 ATGACAGCCTCTTGATGAACCGGAGGAGTCTTTTACCAATTCAAAAATGCCGTGG 60

Qy 140 GCTAAGGGTCGGGTGAGACCTACCTGTGTAGTAGTGAAGAGGGGTGACAGTGCTACA 199
Db 61 GCTAAGGGTCGGGTGAGACCTACCTGTGTAGTAGTGAAGAGGGGTGACAGTGCTACA 120

Qy 200 TCCTTTTCACTGACATTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCGCTC 259
Db 121 TCCTTTTCACTGACATTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCGCTC 180

Qy 260 TTCCTCCGCTACATCTCGGACTCGGACCTAGACCTGGCCGCTGCTACCCGCTCACCTGG 319
Db 181 TTCCTCCGCTACATCTCGGACTCGGACCTAGACCTGGCCGCTGCTACCCGCTCACCTGG 240

Qy 320 TTCACCTCCGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTCCTGCGAGGG 379
Db 241 TTCACCTCCGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTCCTGCGAGGG 300

Qy 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTCTGTGAGGACCGCAAG 360

Qy 440 GCTGAGCCCGAGGGGTGCGCGGCTGCACCGCGCGGGTGCATAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGGTGCGCGGCTGCACCGCGCGGGTGCATAATAGCCATCATGACC 420

Qy 500 TTCAGAATATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAGAACTTTCAA 559
Db 421 TTCAGAATATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAGAACTTTCAA 480

Qy 560 GCCTGGGAGGGGTGCATGAATTCAGTTTCGTTCTCTCCAGACAGCTTCGGCGCATCCTT 619
Db 481 GCCTGGGAGGGGTGCATGAATTCAGTTTCGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540

Qy 620 TTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTTCTTTGGGACTTTG 675
Db 541 TTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTTCTTTGGGACTTTG 596

RESULT 24
AF529835 596 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 8 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION
```

```

QY 500 TTCAAAGATTATTTTACTGCTGGAATFACCTTTTGTAGAAAAACCATGAAAGAACTTTCAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATFACCTTTTGTAGAAAAACCATGAAAGAACTTTCAA 480
QY 560 GCCTGGGAAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCCTT 619
Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCCTT 540
QY 620 TTGCCCCCTGATGAGTTGATGACTTACGAGACGCACTTTCGTACTTTGGGACTTTG 675
Db 541 TTGCCCCCTGATGAGTTGATGACTTACGAGACGCACTTTCGTACTTTGGGACTTTG 596

RESULT 25
AF529837 596 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 10 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529837
VERSION AF529837.1 GI:22297261
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
TITLE Martin,A. and Scharff,M.D.
JOURNAL Somatic hypermutation of the AID transgene in B and non-B cells
REFERENCE Unpublished
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="10"
/cell_type="hybridoma pl-5"
/transgenic
1..596
/source
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95424.1"
/db_xref="GI:22297262"
/translation="MDSLMMNRKKELYOFKNYRWAKGRRETYLCYVVKRRDSATSFSL
DPGYLRNKGCHVELLFYISDWLDLPGRCYRVTFWTSWSPCYDCARHVAFLRGNP
NLSRFTARLFCEDRKAEPGLRHAGVOIAIMTKDFYICWTFVENVHERTFK
AWEGHENSVRSLRRLILPLLYEDDLRDAFTLGL"
BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN
Query Match 19.3%; Score 545; DB 12; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGCCGCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGCCGCTGG 60
QY 140 GCTAAGGGTGGCGTGAGACCTACCTGCTAGTAGTGAAGAGCGGTGACAGTGCTACA 199
Db 61 GCTAAGGGTGGCGTGAGACCTACCTGCTAGTAGTGAAGAGCGGTGACAGTGCTACA 120
QY 200 TCCTTTTCACTGGACTTTCGTTATCTCTCCCAATAAGAACCGGCTGCCACGCTGGAATTCGCTC 259

```

```

Db 121 TCCTTTTCACTGGACTTTCGTTATCTTCGCAATAAGAACCGGCTGCCACGTGGAAATTCCTC 180
QY 260 TTCTCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGGCTGCTACCGGCTGACCTGG 319
Db 181 TTCTCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGGCTGCTACCGGCTGACCTGG 240
QY 320 TTCAACCTCTGGAGCCCTGCTACGACTGTCGCCGACATGTCGCCGACATTCCTGGGAGG 379
Db 241 TTCAACCTCTGGAGCCCTGCTACGACTGTCGCCGACATGTCGCCGACATTCCTGGGAGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGCACTCTCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCTCAAG 360
QY 440 GCTGAGCCCGAGGGGCTGCGGGGCTGCACCGCGCGGGTGCAGAAATAGTCATCATGACC 499
Db 361 GCTGAGCCCGAGGGGCTGCGGGGCTGCACCGCGCGGGTGCAGAAATAGTCATCATGACT 420
QY 500 TTCAAAGATTATTTTACTGCTGGAATFACCTTTTGTAGAAAAACCATGAAAGAACTTTCAA 559
Db 421 TTCAAAGATTATTTTACTGCTGGAATFACCTTTTGTAGAAAAACCATGAAAGAACTTTCAA 480
QY 560 GCCTGGGAAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCCTT 619
Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCCTT 540
QY 620 TTGCCCCCTGATGAGTTGATGACTTACGAGACGCACTTTCGTACTTTGGGACTTTG 675
Db 541 TTGCCCCCTGATGAGTTGATGACTTACGAGACGCACTTTCGTACTTTGGGACTTTG 596

RESULT 26
AF529839 596 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 12 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529839
VERSION AF529839.1 GI:22297265
KEYWORDS house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
TITLE Martin,A. and Scharff,M.D.
JOURNAL Somatic hypermutation of the AID transgene in B and non-B cells
REFERENCE Unpublished
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="12"
/cell_type="hybridoma pl-5"
/transgenic
1..596
/source
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95426.1"
/db_xref="GI:22297266"
/translation="MDSLMMNRKKELYOFKNYRWAKGRRETYLCYVVKRRDSATSFSL
DPGYLRNKGCHVELLFYISDWLDLPGRCYRVTFWTSWSPCYDCARHVAFLRGNP

```

BASE COUNT 128 a 163 c 155 g 150 t
ORIGIN
NLSRIETARLYFCEDRKAETEGRLRLHRAGVQIAIMTFKDYFYCWNTEFVENHETFK
AWEGJHNSVRLSRILLPLLYEVDLDRFAFTLGL"

Query Match 19.3%; Score 545; DB 12; Length 596;
Best Local Similarity 99.8%; Pred. No. 9.8e-273;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAAATCGCGCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAAATCGCGCTGG 60
QY 140 GCTAAGGGTCGCGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 199
Db 61 GCTAAGGGTCGCGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120
QY 200 TCTTTTCTACTGGACTTTGGTTATCTTCGCAATAGACGGTGCACAGTGAATTTGCTC 259
Db 121 TCTTTTCTACTGGACTTTGGTTATCTTCGCAATAGACGGTGCACAGTGAATTTGCTC 180
QY 260 TTCTTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGGCTACCTGG 319
Db 181 TTCTTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGGCTACCTGG 240
QY 320 TTACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGGGAGGG 379
Db 241 TTACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGGGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGGCTTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGGCTTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCGAGGGCTGCGCGGCTGCACCGCGGGGTGCAATAGCCATCATGACC 499
Db 361 GCTGAGCCGAGGGCTGCGCGGCTGCACCGCGGGGTGCAATAGCCATCATGACC 420
QY 500 TTCAAGATTAATTTTACTGCTGGAATCTTTTGTAGAAACCATGAAGAACTTTTCAA 559
Db 421 TTCAAGATTAATTTTACTGCTGGAATCTTTTGTAGAAACCATGAAGAACTTTTCAA 480
QY 560 GCTGGGAAGGCTGCATGAATAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTTT 619
Db 481 GCTGGGAAGGCTGCATGAATAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTTT 540
QY 620 TTGCCCCCTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 675
Db 541 TTGCCCCCTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 596

RESULT 27
AF529834
LOCUS Mus musculus clone 7 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529834
VERSION AF529834.1 GI:22297255
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 595)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 595)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
1. .595 Location/Qualifiers
source

source
1. .595
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="7"
/cell_type="hybridoma P1-5"
/transgenic
gene
1. .>595
/organism="Homo sapiens"
/db_xref="taxon:9606"
CDS
1. .>595
/gene="AID"
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AA095421.1"
/db_xref="GI:22297256"
/translation="MDSLMMRRKFLYQKVRWAKGRRETYLCVIVKRRSATSFSLS
DFGLRNKNGCHVELLFURYISDMDPGRYRVWTFWSPCYDCARHVAFLRGNP
NLSRIETARLYFCEDRKAETEGRLRLHRAGVQIAIMTFKDYFYCWNTEFVENHETFK
AWEGJHNSVRLSRILLPLLYEVDLDRFAFTLGL"

BASE COUNT 127 a 163 c 154 g 151 t
ORIGIN
Query Match 19.3%; Score 544; DB 12; Length 595;
Best Local Similarity 99.8%; Pred. No. 3.3e-272;
Matches 594; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAAATCGCGCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTTCTTTACCAATTCAAAAATCGCGCTGG 60
QY 140 GCTAAGGGTCGCGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 199
Db 61 GCTAAGGGTCGCGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120
QY 200 TCTTTTCTACTGGACTTTGGTTATCTTCGCAATAGACGGTGCACAGTGAATTTGCTC 259
Db 121 TCTTTTCTACTGGACTTTGGTTATCTTCGCAATAGACGGTGCACAGTGAATTTGCTC 180
QY 260 TTCTTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGGCTACCTGG 319
Db 181 TTCTTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGGCTACCTGG 240
QY 320 TTACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGGGAGGG 379
Db 241 TTACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGGGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGGCTTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGGCTTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCGAGGGCTGCGCGGCTGCACCGCGGGGTGCAATAGCCATCATGACC 499
Db 361 GCTGAGCCGAGGGCTGCGCGGCTGCACCGCGGGGTGCAATAGCCATCATGACC 420
QY 500 TTCAAGATTAATTTTACTGCTGGAATCTTTTGTAGAAACCATGAAGAACTTTTCAA 559
Db 421 TTCAAGATTAATTTTACTGCTGGAATCTTTTGTAGAAACCATGAAGAACTTTTCAA 480
QY 560 GCTGGGAAGGCTGCATGAATAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTTT 619
Db 481 GCTGGGAAGGCTGCATGAATAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTTT 540
QY 620 TTGCCCCCTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTT 674
Db 541 TTGCCCCCTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTT 595

RESULT 28
AF529841
LOCUS Cricetus griseus clone 1 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
597 bp mRNA linear SYN 19-AUG-2002

```

ACCESSION AF529841 GI:22297269
VERSION AF529841.1
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
    source
        1..597
            /organism="Cricetulus griseus"
            /db_xref="taxon:10029"
            /clone="1"
            /cell_line="CHO"
            /transgenic
            1..597
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            1..597
            /gene="AID"
            1..597
            /gene="AID"
            /codon_start=1
            /product="AID"
            /protein_id="AAM95428.1"
            /db_xref="GI:22297270"
            /translation="MDSLMMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATFSFL
            DFGYLRNKGCHVELLFLRYISDWLDPGRCYRVTWFTSWSPCYDCAHVADFLRGNP
            NLSRIPTARLYFCEDRKAEPGLRLHRAQVQIAIMTKDYFCWNTFVENHRTFFK
            AWEGLHENSVRSLRRLILLPLYEVDLDAFTWGR"
        129 a 164 c 157 g 147 t

BASE COUNT
ORIGIN
Query Match 19.0%; Score 535; DB 12; Length 597;
Best Local Similarity 99.8%; Pred. No. 1.6e-267;
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 60

QY 140 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 199
Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 120

QY 200 TCCCTTTTACTGGACTTTGGTTATCTTCCCAATAAGAACGGTGGCCACGTGGAATGGTC 259
Db 121 TCCCTTTTACTGGACTTTGGTTATCTTCCCAATAAGAACGGTGGCCACGTGGAATGGTC 180

QY 260 TTCCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACTGG 319
Db 181 TTCCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACTGG 240

QY 320 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTCTTGGGAGG 379
Db 241 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTCTTGGGAGG 300

QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCTCTACTTCTGTGAGGACCGCAAG 360

QY 440 GCTGAGCCCGAGGGCTGGCGGCTGACACCGCGGGGGTGC AATAGCCATCATGACC 499

```

```

Db 361 GCTGAGCCCGAGGGCTGCGCGGCTGCACCGCGGGTGC AAAATAGCATCATGATC 420
QY 500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTCAA 559
Db 421 ATCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTCAA 480
QY 560 GCCTGGGAAGGCGCTGCATGAAAATTCAGTTCTGCTCTCCAGACAGCTTCGGTGCAT 619
Db 481 GCCTGGGAAGGCGCTGCATGAAAATTCAGTTCTGCTCTCCAGACAGCTTCGGTGCAT 540
QY 620 TTGCCCCCTATGAGGTTGATGACATTACGAGCGCATTTTCGTACTT 665
Db 541 TTGCCCCCTATGAGGTTGATGACATTACGAGCGCATTTTCGTACTT 586

RESULT 29
AF529843 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 3 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529843
VERSION AF529843.1 GI:22297273
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
    source
        1..597
            /organism="Cricetulus griseus"
            /db_xref="taxon:10029"
            /clone="3"
            /cell_line="CHO"
            /transgenic
            1..597
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            1..597
            /gene="AID"
            1..597
            /gene="AID"
            /codon_start=1
            /product="AID"
            /protein_id="AAM95430.1"
            /db_xref="GI:22297274"
            /translation="MDSLMMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATFSFL
            DFGYLRNKGCHVELLFLRYISDWLDPGRCYRVTWFTSWSPCYDCAHVADFLRGNP
            NLSRIPTARLYFCEDRKAEPGLRLHRAQVQIAIMTKDYFCWNTFVENHRTFFK
            AWEGLHENSVRSLRRLILLPLYEVDLDAFTWGR"
        129 a 164 c 156 g 148 t

BASE COUNT
ORIGIN
Query Match 19.0%; Score 535; DB 12; Length 597;
Best Local Similarity 99.8%; Pred. No. 1.6e-267;
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG 60

QY 140 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 199

```

Db 61 GCTAAGGCTCGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTCTACA 120
Qy 200 TCCCTTTTTCACCTGGACTTTGGTTATCTTCGCAATAGAAGCGCTGCCAGTGAATTCGCTC 259
Db 121 TCCCTTTTTCACCTGGACTTTGGTTATCTTCGCAATAGAAGCGCTGCCAGTGAATTCGCTC 180
Qy 260 TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGGCTCACCTGG 319
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGGCTCACCTGG 240
Qy 320 TTCACCTCTCGAGCCCTCTAGACACTGTGCGCGACATGTGGCGGACTTCTTCGGAGGG 379
Db 241 TTCACCTCTCGAGCCCTCTAGACACTGTGCGCGACATGTGGCGGACTTCTTCGGAGGG 300
Qy 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCCCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCCCTCTACTTCTGTGAGGACCGCAAG 360
Qy 440 GCTGAGCCCGAGGGCTCGCGCGCTGCACCGCGCGGCTGCAAAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGCTCGCGCGCTGCACCGCGCGGCTGCAAAATAGCCATCATGACC 420
Qy 500 TTCAAGATTTATTTTACTGCTGGATACCTTTTGTAGAAACCATGAAGAACTTTTCAA 559
Db 421 TTCAAGATTTATTTTACTGCTGGATACCTTTTGTAGAAACCATGAAGAACTTTTCAA 480
Qy 560 GCCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 619
Db 481 GCCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Qy 620 TTGCCCTCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTT 665
Db 541 TTGCCCTCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTT 586

RESULT 30
AF529844
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
source
gene
CDS

AF529844
Cricetulus griseus clone 4 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
AF529844.1 GI:22297275
Chinese hamster.
Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
1 (bases 1 to 597)
Martin,A. and Scharff,M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
Unpublished
2 (bases 1 to 597)
Martin,A. and Scharff,M.D.
Direct Submission
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
Location/Qualifiers
1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="4"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/codon_start=1

/product="AID"
/protein_id="AA95431.1"
/db_xref="GI:22297276"
/translation="MDSLLNNRRKFLYQFNVRWAKGRRETYLCYVYVVRSDATSFSLS
DFGLRNKNGCHEVELLEFLRYISMDLDPGRCYRWFVTSWSPCYDCARHVAADFURGNP
NLSURITFARLYFCEDRKAEPGRULRHRAGVQIAIMTFKDYFCWNTFVENHERTEK
AWEGHNSVRLSRQLRILLPLVEVDLRDAFRTWGR"
BASE COUNT 128 a 163 c 158 g 148 t
ORIGIN
Query Match 19.0%; Score 535; DB 12; Length 597;
Best Local Similarity 99.8%; Pred. No. 1.6e-267;
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 80 ATGACAGCCCTTCATGATCAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGACAGCCCTTCATGATCAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCTGG 60
Qy 140 GCTAAGGCTCGCGGTGAGACCTACCTGTGTCTACGTAGTGAAGAGCGCTGACAGTCTACA 199
Db 61 GCTAAGGCTCGCGGTGAGACCTACCTGTGTCTACGTAGTGAAGAGCGCTGACAGTCTACA 120
Qy 200 TCCCTTTTTCACCTGGACTTTGGTTATCTTCGCAATAGAAGCGCTGCCAGTGAATTCGCTC 259
Db 121 TCCCTTTTTCACCTGGACTTTGGTTATCTTCGCAATAGAAGCGCTGCCAGTGAATTCGCTC 180
Qy 260 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCCCTGCACCGCGCTCTACCTGCTCACCTGG 319
Db 181 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCCCTGCTACCGCGCTGCTACCGCTCACCTGG 240
Qy 320 TTCAACCTCTGTGAGCCCTCTGCTACGACTGTGCCCGGACATGTGGCGGACTTCTTCGAGGG 379
Db 241 TTCAACCTCTGTGAGCCCTCTGCTACGACTGTGCCCGGACATGTGGCGGACTTCTTCGAGGG 300
Qy 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGGCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGGCTCTACTTCTGTGAGGACCGCAAG 360
Qy 440 GCTGAGCCCGAGGGCTCGCGCGCTGCACCGCGCGGCTGCAAAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGCTCGCGCGCTGCACCGCGCGGCTGCAAAATAGCCATCATGACC 420
Qy 500 TTCAAGATTTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAGAACTTTTCAA 559
Db 421 TTCAAGATTTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAGAACTTTTCAA 480
Qy 560 GCCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGGATCCTT 619
Db 481 GCCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGGATCCTT 540
Qy 620 TTGCCCTCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTT 665
Db 541 TTGCCCTCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTT 586
RESULT 31
AF529845
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

AF529845
Cricetulus griseus clone 5 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
AF529845.1 GI:22297277
Chinese hamster.
Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
1 (bases 1 to 597)
Martin,A. and Scharff,M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
Unpublished

QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCTCTACTTCTGTGAGGACCGCAAG 439
|||||
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCTCTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCCGAGGGCTCGCGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGCTCGCGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
QY 500 TTCAAGATATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAGAACTTTCAAA 559
Db 421 TTCAAGATATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAGAACTTTCAAA 480
QY 560 GCCTGGGAAGGGTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 619
Db 481 GCCTGGGAAGGGTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 540
QY 620 TTGCCCCCTGTAGAGTTGATGACTTACGAGAGCGCATTTTCGTACTT 665
Db 541 TTGCCCCCTGTAGAGTTGATGACTTACGAGAGCGCATTTTCGTACTT 586

RESULT 33
AF529849
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AF529849 597 bp mRNA linear SYN 19-AUG-2002
Cricetulus griseus clone 9 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
AF529849
AF529849.1 GI:22297285
Chinese hamster.
Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
1 (bases 1 to 597)
Martin, A. and Scharff, M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
Unpublished
2 (bases 1 to 597)
Martin, A. and Scharff, M.D.
Direct Submission
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
Location/Qualifiers
1. .597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="9"
/cell_line="CHO"
/transgenic

source

gene

CDS

BASE COUNT 127 a 164 c 158 g 148 t

ORIGIN

Query Match 19.0%; Score 535; DB 12; Length 597;
Best Local Similarity 99.8%; Pred. No. 1.6e-267;
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 80 ATGCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTTACCAATTCAAAATGTCCGCTGG 139
|||||
Db 1 ATGCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTTACCAATTCAAAATGTCCGCTGG 60
QY 140 GCTAAGGGTGGCGGTGAGACCTACCTGTGCTAGTGAAGAGGGGTGACAGTGTCTACA 199
Db 61 GCTAAGGGTGGCGGTGAGACCTACCTGTGCTAGTGAAGAGGGGTGACAGTGTCTACA 120
QY 200 TCCCTTTTCACTGACATTTGGTTATCTTCCAAATGAAGACGGCTGCCACGTGGAAATGCTC 259
Db 121 TCCCTTTTCACTGACATTTGGTTATCTTCCAAATGAAGACGGCTGCCACGTGGAAATGCTC 180
QY 260 TTCTCTCCGTACATCTCGGACTGGGACTAGACCCCTGGCGCTGCTACCCGCTCACCTGG 319
Db 181 TTCTCTCCGTACATCTCGGACTGGGACTAGACCCCTGGCGCTGCTACCCGCTCACCTGG 240
QY 320 TTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTTCGCGAGGG 379
Db 241 TTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTTCGCGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCTCTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCCGAGGGTTCGGCGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGTTCGGCGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
QY 500 TTCAAGATATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAGAACTTTCAAA 559
Db 421 TTCAAGATATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAGAACTTTCAAA 480
QY 560 GCCTGGGAAGGGTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 619
Db 481 GCCTGGGAAGGGTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 540
QY 620 TTGCCCCCTGTAGAGTTGATGACTTACGAGAGCGCATTTTCGTACTT 665
Db 541 TTGCCCCCTGTAGAGTTGATGACTTACGAGAGCGCATTTTCGTACTT 586

RESULT 34
AF529850
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AF529850 597 bp mRNA linear SYN 19-AUG-2002
Cricetulus griseus clone 10 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
AF529850
AF529850.1 GI:22297287
Chinese hamster.
Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
1 (bases 1 to 597)
Martin, A. and Scharff, M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
Unpublished
2 (bases 1 to 597)
Martin, A. and Scharff, M.D.
Direct Submission
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
Location/Qualifiers
1. .597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="10"
/cell_line="CHO"
/transgenic

source

gene

CDS

BASE COUNT 127 a 164 c 158 g 148 t

ORIGIN

Query Match 19.0%; Score 535; DB 12; Length 597;
Best Local Similarity 99.8%; Pred. No. 1.6e-267;
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

gene	/organism="Homo sapiens"									
	/db_xref="taxon:9606"									
	1..597									
	/gene="AID"									
CDS	1..597									
	/gene="AID"									
	/codon_start=1									
	/product="AID"									
	/protein_id="AAM95437.1"									
	/db_xref="GI:22297288"									
	/translation="MDSLNNRRKFLYQFNVRWAKGRRETYLCVYVVKRRDSATSFSL DFGYLRNKGCHVELFLRYISDWLDLPVRCYRVTFWTSWSPCYDCARHVADFLRGNP NLSLRFTARLYFCEDRKAEPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHERTEFK ANEGLHENSVRSLRQLRILLPLYEVDLDRDAFTWGR"									
	BASE COUNT	128 a	164 c	156 g	149 t					
ORIGIN										
Query Match	19.0%; Score 535; DB 12; Length 597;									
Best Local Similarity	99.8%; Pred. No. 1.6e-267;									
Matches	585;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;	
QY	80	ATGCACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGCCGTGG	139							
Db	1	ATGCACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGCCGTGG	60							
QY	140	GCTAAGGGTCGGCGTCGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA	199							
Db	61	GCTAAGGGTCGGCGTCGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA	120							
QY	200	TCCTTTTACTGACATCTGGTTATCTTCGCAATAAGAACGGTGCACGTGGAATTGCTC	259							
Db	121	TCCTTTTACTGACATCTGGTTATCTTCGCAATAAGAACGGTGCACGTGGAATTGCTC	180							
QY	260	TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCGCTGCTACCGGTCACTCTGG	319							
Db	181	TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGCTCCGCTGCTACCGGTCACTCTGG	240							
QY	320	TTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGCAGATTTCTGCGGAGGG	379							
Db	241	TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGCAGATTTCTGCGGAGGG	300							
QY	380	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	439							
Db	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	360							
QY	440	GCTGAGCCGAGGGGTGGCGCGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC	499							
Db	361	GCTGAGCCGAGGGGTGGCGCGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC	420							
QY	500	TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAAGCTTTTCAA	559							
Db	421	TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAAGCTTTTCAA	480							
QY	560	GCCTGGGAAGGGCTGCATGAAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCCTT	619							
Db	481	GCCTGGGAAGGGCTGCATGAAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCCTT	540							
QY	620	TTGCCCTCTGATGAGGTTGATGACTTTACGAGACGCATTTTCGTACTT	665							
Db	541	TTGCCCTCTGATGAGGTTGATGACTTTACGAGACGCATTTTCGTACTT	586							
RESULT 35										
AF529851										
LOCUS										
DEFINITION	AF529851 Cricetulus griseus clone 11 transgenic Homo sapiens AID (AID) mRNA, complete cds.									
ACCESSION	AF529851									
VERSION	AF529851.1 GI:22297289									
KEYWORDS										
SOURCE	Chinese hamster.									
ORGANISM	Cricetulus griseus									
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									

		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetidae;									
		Cricetulus.									
REFERENCE	1	(bases 1 to 597)									
AUTHORS	Martin,A. and Scharff,M.D.										
TITLE	Somatic hypermutation of the AID transgene in B cells and non-B										
JOURNAL	cells										
REFERENCE	2	(bases 1 to 597)									
AUTHORS	Martin,A. and Scharff,M.D.										
TITLE	Direct Submission										
JOURNAL	Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of										
	Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA										
FEATURES	Location/Qualifiers										
Source	1..597	/organism:"Cricetulus griseus"									
		/db_xref:"taxon:10029"									
		/clone="11"									
		/cell_line:"CHO"									
		/transgenic									
source	1..597	/organism:"Homo sapiens"									
		/db_xref:"taxon:9606"									
gene	1..597	/gene="AID"									
CDS	1..597	/gene="AID"									
		/codon_start=1									
		/product:"AID"									
		/protein_id:"AAM95438.1"									
		/db_xref:"GI:22297290"									
		/translation:"MDSLNNRRKFLYQFNVRWAKGRRETYLYVVKRRUSATSFSL									
		DFGYLRNKGCHVELLFURYISDWLDPQRCYRVTWFTSWSPCYDCARIVADFLRGNP									
		NLSLRIFARLYFCEDRKAEPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHERTEFKP									
		AWEGIHENSVRPLPROLRLILLPLLEYEVDLDRDAFRTWGR"									
BASE COUNT	128 a	165 c	157 g	147 t							
ORIGIN											
Query Match	19.0%; Score 535; DB 12; Length 597;										
Best Local Similarity	99.8%; Pred. No. 1.6e-267;										
Matches	585;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;		
QY	80	ATGGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGCCCTGG									139
Db	1	ATGGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGCCCTGG									60
QY	140	GCTAAGGGTCGGCGTCGAGACCTACCTGTGCTAGTGAAGAGCGCTGACAGTGCTACA									199
Db	61	GCTAAGGGTCGGCGTCGAGACCTACCTGTGCTAGTGAAGAGCGCTGACAGTGCTACA									120
QY	200	TCCTTTTACTGACATCTGGTTATCTTCGCAATAAGAACGGTGCACGTGGAATTCGTC									259
Db	121	TCCTTTTACTGACATCTGGTTATCTTCGCAATAAGAACGGTGCACGTGGAATTCGTC									180
QY	260	TTCTCCGCTACATCTCGGACCTGGACCTAGACCTGGCGGCTGCTACCGCTCACCTGG									319
Db	181	TTCTCCGCTACATCTCGGACCTGGACCTAGACCTGGCGGCTGCTACCGCTCACCTGG									240
QY	320	TTACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTTGGCAGGG									379
Db	241	TTACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTTGGCAGGG									300
QY	380	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGAAG									439
Db	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGAAG									460
QY	440	GCTGAGCCGAGGGCTCGGGCGGCTGGACCGCGCGGGGTGCAATAGCTATCATGATGC									499
Db	361	GCTGAGCCGAGGGCTCGGGCGGCTGGACCGCGCGGGGTGCAATAGCTATCATGATGC									420
QY	500	TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATCTTTCAAA									559
Db	421	TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATCTTTCAAA									480


```
QY 560 GCCTGGGAAGGCGTCATGAAATTCAGTTCTCTCTCCAGACAGCTTCGGCGCATCCTT 619
|||||
Db 481 GCCTGGGAAGGCGTCATGAAATTCAGTTCTCTCTCCAGACAGCTTCGGCGCATCCTT 540

QY 620 TTGCCCCGTATGAGTTGATGACTTACGAGACGATTTCTGACTT 665
|||||
Db 541 TTGCCCCGTATGAGTTGATGACTTACGAGACGATTTCTGACTT 586

RESULT 36
AF529852 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 12 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529852
VERSION AF529852.1 GI:22297291
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="12"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95439.1"
/db_xref="GI:22297292"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFRLYISDMDLPGRCYRVTFWTSWSPCYDCARHVAFLRGNP
NLSLRIFTARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTFVENHRTFK
AWEGHENSVRLSRQLRLPLLYEVDLDRDAFTWGR"
BASE COUNT 129 a 164 c 156 g 148 t
ORIGIN

Query Match 19.0%; Score 535; DB 12; Length 597;
Best Local Similarity 99.8%; Pred. No. 1.6e-267;
Matches 585; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 80 ATGACACGCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGGCTGG 139
|||||
Db 1 ATGACACGCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGGCTGG 60

QY 140 GCTAAGGTCGGCGTGAGACCTACTGCTGCTAGCTAGTGAAGCGCTGACAGTGCTACA 199
|||||
Db 61 GCTAAGGTCGGCGTGAGACCTACTGCTGCTAGCTAGTGAAGCGCTGACAGTGCTACA 120

QY 200 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGCTGCGACGTGGAATTGCTC 259
|||||
Db 121 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGCTGCGACGTGGAATTGCTC 180
```

```
QY 260 TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGSCCGCTGCTACCGCGTCACCTGG 319
|||||
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGSCCGCTGCTACCGCGTCACCTGG 240

QY 320 TTCACCTCTCGAGCCCTGCTAGACTGTGCCGACATGTGGCGGACTTCTTCTGGGAGG 379
|||||
Db 241 TTCACCTCTCGAGCCCTGCTAGACTGTGCCGACATGTGGCGGACTTCTTCTGGGAGG 300

QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAG 439
|||||
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360

QY 440 GCTGAGCCCGAGGGCTGCGCGCTGCACCGCCCGGGGTGCAAAATAGCCATCATGACC 499
|||||
Db 361 GCTGAGCCCGAGGGACTGCGCGCTGCACCGCCCGGGGTGCAAAATAGCCATCATGACC 420

QY 500 TTCAAAGATTATTTTACTGCTGGAATCTTTCTAGAAAACCATGAAGAATTTTCAA 559
|||||
Db 421 TTCAAAGATTATTTTACTGCTGGAATCTTTCTAGAAAACCATGAAGAATTTTCAA 480

QY 560 GCCTGGGAAGGCGTCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 619
|||||
Db 481 GCCTGGGAAGGCGTCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540

QY 620 TTGCCCCGTATGAGTTGATGACTTACGAGACGATTTCTGACTT 665
|||||
Db 541 TTGCCCCGTATGAGTTGATGACTTACGAGACGATTTCTGACTT 586

RESULT 37
AF529853 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 13 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529853
VERSION AF529853.1 GI:22297293
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="13"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95440.1"
/db_xref="GI:22297294"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFRLYISDMDLPGRCYRVTFWTSWSPCYDCARHVAFLRGNP
NLSLRIFTARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFYWNTFVENHRTFK
```

BASE COUNT		129 a	164 c	156 g	148 t
ORIGIN		AWEGLHNSVRLSRQLRILLPLVEVDLDRDAFRTWGR"			
Query Match		19.0%;	Score 535;	DB 12;	Length 597;
Best Local Similarity		99.8%;	Pred. No. 1.6e-267;		
Matches 585;		Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	80	ATGGACAGCCTCTTCATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCGTGG	139		
Db	1	ATGGACAGCCTCTTCATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCGTGG	60		
Qy	140	GCTAAGGTCGCGTGAGACCTACCTGTGCTAGTGTAGTGAAGAGCGGTGACAGTGTCTACA	199		
Db	61	GCTAAGGTCGCGTGAGACCTACCTGTGCTAGTGTAGTGAAGAGCGGTGACAGTGTCTACA	120		
Qy	200	TCCTTTTTCACGTGAGCTTTGGTTATCTTCGCAATAGAACGGGTGCCAGTGAATTTGCTC	259		
Db	121	TCCTTTTTCACGTGAGCTTTGGTTATCTTCGCAATAGAACGGGTGCCAGTGAATTTGCTC	180		
Qy	260	TTCTCTCCGCTACATCTCGGACTTGGGACCTAGACCTGTGCGCGCTGCTACCGGTACCTGG	319		
Db	181	TTCTCTCCGCTACATCTCGGACTTGGGACCTAGACCTGTGCGCGCTGCTACCGGTACCTGG	240		
Qy	320	TTGACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTTTCGCGAGGG	379		
Db	241	TTGACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTTTCGCGAGGG	300		
Qy	380	AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	439		
Db	301	AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	160		
Qy	440	GCTGAGCCCGAGGGCTGCGGGCGCTGCACCGCGCGGGTGCAAAATAGCATCATGACC	499		
Db	361	GCTGAGCCCGAGGGCTGCGGGCGCTGCACCGCGCGGGTGCAAAATAGCATCATGACC	420		
Qy	500	TTCAAAAGATTATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAGAATTTTCAAA	559		
Db	421	TTCAAAAGATTATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAGAATTTTCAAA	480		
Qy	560	GCCTGGGAGGGCTGCATGAAAATTCAGTTTCGCTCTCCAGACAGCTTCGGGCGATCTT	619		
Db	481	GCCTGGGAGGGCTGCATGAAAATTCAGTTTCGCTCTCCAGACAGCTTCGGGCGATCTT	540		
Qy	620	TTGCCCTCTGTATGAGGTTGATGACTTACGAGACGCATTTTCGTACTT	665		
Db	541	TTGCCCTCTGTATGAGGTTGATGACTTACGAGACGCATTTTCGTACTT	586		
RESULT 38					
AF529854					
LOCUS		597 bp	mRNA	linear	SYN 19-AUG-2002
DEFINITION		Cricetulus griseus clone 14 transgenic Homo sapiens AID (AID) mRNA, complete cds.			
ACCESSION		AF529854			
VERSION		AF529854.1 GI:22297295			
KEYWORDS		Chinese hamster.			
SOURCE		Cricetulus griseus			
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.			
REFERENCE		1 (bases 1 to 597)			
AUTHORS		Martin,A. and Scharff,M.D.			
TITLE		Somatic hypermutation of the AID transgene in B cells and non-B cells			
JOURNAL		unpublished			
REFERENCE		2 (bases 1 to 597)			
AUTHORS		Martin,A. and Scharff,M.D.			
TITLE		Direct Submission			
JOURNAL		Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA			
FEATURES		Location/Qualifiers			
BASE COUNT		128 a	163 c	158 g	148 t
ORIGIN					
Query Match		19.0%;	Score 535;	DB 12;	Length 597;
Best Local Similarity		99.8%;	Pred. No. 1.6e-267;		
Matches 585;		Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	80	ATGGACAGCCTCTTCATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCGTGG	139		
Db	1	ATGGACAGCCTCTTCATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCGTGG	60		
Qy	140	GCTAAGGTCGCGTGAGACCTACCTGTGCTAGTGTAGTGAAGAGCGGTGACAGTGTCTACA	199		
Db	61	GCTAAGGTCGCGTGAGACCTACCTGTGCTAGTGTAGTGAAGAGCGGTGACAGTGTCTACA	120		
Qy	200	TCCTTTTTCACGTGAGCTTTGGTTATCTTCGCAATAGAACGGGTGCCAGTGAATTTGCTC	259		
Db	121	TCCTTTTTCACGTGAGCTTTGGTTATCTTCGCAATAGAACGGGTGCCAGTGAATTTGCTC	180		
Qy	260	TTCTCTCCGCTACATCTCGGACTTGGGACCTAGACCTGTGCGCGCTGCTACCGGTACCTGG	319		
Db	181	TTCTCTCCGCTACATCTCGGACTTGGGACCTAGACCTGTGCGCGCTGCTACCGGTACCTGG	240		
Qy	320	TTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGAGCTTTTTCGCGAGGG	379		
Db	241	TTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGAGCTTTTTCGCGAGGG	300		
Qy	380	AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	439		
Db	301	AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	360		
Qy	440	GCTGAGCCCGAGGGCTGCGGGCGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACC	499		
Db	361	GCTGAGCCCGAGGGCTGCGGGCGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACC	420		
Qy	500	TTCAAGATTATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAGAATTTTCAAA	559		
Db	421	TTCAAGATTATTTTACTGCTGGAATACATTTTGTAGAAAACCATGAAGAATTTTCAAA	480		
Qy	560	GCCTGGGAGGGCTGCATGAAAATTCAGTTTCGCTCTCCAGACAGCTTTCGGCGCATCCTT	619		
Db	481	GCCTGGGAGGGCTGCATGAAAATTCAGTTTCGCTCTCCAGACAGCTTTCGGCGCATCCTT	540		
Qy	620	TTGCCCTCTGTATGAGGTTGATGACTTACGAGACGCATTTTCGTACTT	665		
Db	541	TTGCCCTCTGTATGAGGTTGATGACTTACGAGACGCATTTTCGTACTT	586		
RESULT 39					
AF529855					
LOCUS		591 bp	mRNA	linear	SYN 19-AUG-2002
DEFINITION		Cricetulus griseus clone 15 transgenic Homo sapiens AID (AID) mRNA.			

```
complete cds.
AF529855.1 GI:22297297
VERSION
KEYWORDS
SOURCE
ORGANISM
Chinese hamster.
Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE
1 (bases 1 to 591)
Martin, A. and Scharff, M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 591)
Martin, A. and Scharff, M.D.
Direct Submission
TITLE
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
JOURNAL
Location/Qualifiers
FEATURES
source
1..591
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="15"
/cell_line="CHO"
/transgenic
1..591
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..591
/gene="AID"
<1..591
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95442.1"
/db_xref="GI:22297298"
/translaton="SLLMNRKFLYQFKNVRWAKRGCTYLCYVVKRRDSATSFSLDF
GYLRNKGCHVELLFRIYISDWLDPGRCYRVTFWTSWPCYDCARHVAFLRGNPL
SLRFTFARLYFCEDRAEPEGLRLHRAQVOIAIMTFKDYFCWNTFVENHRTFAW
EGLHNSVRLSRLRLTLLPLYEVDLDRDAFTWGR"
BASE COUNT 126 a 162 c 155 g 148 t
ORIGIN
Query Match 18.8%; Score 529; DB 12; Length 591;
Best Local Similarity 99.8%; Pred. No. 2.2e-264;
Matches 579; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 86 AGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAATGTCGCTGGGCTAAG 145
Db 1 AGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAATGTCGCTGGGCTAAG 60
Qy 146 GTCGCGCTGAGACCTACCTGTGCTAGTGTAGTGAAGAGCGGTGACAGTGTACATCCTTT 205
Db 61 GTCGCTGAGACCTACCTGTGCTAGTGTAGTGAAGAGCGGTGACAGTGTACATCCTTT 120
Qy 206 TCACTGGACTTTGGTTATCTTCGCAATAAAGACGGCTGCCACGTGGGAATTCCTTCCTC 265
Db 121 TCACTGGACTTTGGTTATCTTCGCAATAAAGACGGCTGCCACGTGGGAATTCCTTCCTC 180
Qy 266 CGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCTACCTGGTTCAAC 325
Db 181 CGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCTACCTGGTTCAAC 240
Qy 326 TCCTGGAGCCCTCTACGACTGTGCCGACATGTGGCCGACTTCTTCGAGGGAACCCC 385
Db 241 TCCTGGAGCCCTCTACGACTGTGCCGACATGTGGCCGACTTCTTCGAGGGAACCCC 300
Qy 386 AACCTCAGTCTGAGGATCTTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAG 445
Db 301 AACCTCAGTCTGAGGATCTTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAG 360
Qy 446 CCCGAGGGGCTGCGGGGCTGACACCGCGCGGGGTGCAAAATAGCCATCATGACCTTCAAA 505
```

```
|||||
361 CCGAGGGCTGCGGGCTGCACCGCGGGTGCAATAGCCATCATGACTTCAAA 420
Qy 506 GATTATTTTACTGCTGGAACTACTTTGTAGAAAACCATGAAAGAACTTTTCAAGCCTGG 565
Db 421 GATTATTTTACTGCTGGAACTACTTTGTAGAAAACCATGAAAGAACTTTTCAAGCCTGG 480
Qy 566 GAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATTCCTTTTGCCC 625
Db 481 GAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATTCCTTTTGCCC 540
Qy 626 CTGTATGAGTTGATGACTTACGAGAGCGCATTTTCGTACTT 665
Db 541 CTGTATGAGTTGATGACTTACGAGAGCGCATTTTCGTACTT 580

RESULT 40
AF529840 577 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 13 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529840
VERSION AF529840.1 GI:22297267
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 577)
Martin, A. and Scharff, M.D.
Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 577)
Martin, A. and Scharff, M.D.
Direct Submission
TITLE Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
JOURNAL Location/Qualifiers
FEATURES
source
1..577
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="13"
/cell_type="hybridoma pl-5"
/transgenic
1..577
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..577
/gene="AID"
<1..577
/gene="AID"
/codon_start=3
/product="AID"
/protein_id="AAM95427.1"
/db_xref="GI:22297268"
/translaton="RRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRN
KNGCHVELLFRIYISDWLDPGRCYRVTFWTSWPCYDCARHVAFLRGNPLSLRIF
TARLYFCEDRAEPEGLRLHRAQVOIAIMTFKDYFCWNTFVENHRTFKAWEGLHE
NSVRLSQRLLRLPLYEVDLDRDAFTLGL"
BASE COUNT 122 a 159 c 150 g 146 t
ORIGIN
Query Match 18.7%; Score 526; DB 12; Length 577;
Best Local Similarity 99.8%; Pred. No. 8e-263;
Matches 576; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 99 ACCGGAGGAAGTTTCTTTACCAATTCAAAATGTCGCTGGGCTAAGGTCGGCGTGAGA 158
Db 1 ACCGGAGGAAGTTTCTTTACCAATTCAAAATGTCGCTGGGCTAAGGTCGGCGTGAGA 60
Qy 159 CCTACCTGTCTACGTAGTGAAGAGCGCTGACAGTGTACATCCTTTTCTACTGACCTTG 218
Db 61 CCTACCTGTCTACGTAGTGAAGAGCGCTGACAGTGTACATCCTTTTCTACTGACCTTG 120
```

```
QY 219 GTTATCTTGCATATAGAACGGCTGCCACGTGGAATGCTCTTCCCTCCGTACATCTCGG 278
D 121 GTTATCTTGCATATAGAACGGCTGCCACGTGGAATGCTCTTCCCTCCGTACATCTCGG 180
QY 279 ACTGGGACCTAGACCCCTGGCCGTGTACCGCTCACCCTGTTTCACTCTCTGGAGCCCT 338
D 181 ACTGGGACCTAGACCCCTGGCCGTGTACCGCTCACCCTGTTTCACTCTCTGGAGCCCT 240
QY 339 GCTACGACTGTGCCGACATGTGGCGACTTCTCGGAGGAACTTCTCGGAGGAACTTCACTCTG 398
D 241 GCTACGACTGTGCCGACATGTGGCGACTTCTCGGAGGAACTTCTCGGAGGAACTTCACTCTG 300
QY 399 GGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGC 458
D 301 GGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGC 360
QY 459 GCGGGCTGACCGCGCGGGTGCCTTCTGAGGACCGCAAGGCTGAGCCGAGGGGCTGC 518
D 361 GCGGGCTGACCGCGCGGGTGCCTTCTGAGGACCGCAAGGCTGAGCCGAGGGGCTGC 420
QY 519 GCTGGAATCTTTGTAGAAACCATGAAAGAACTTTTCAAGCCTGGGAGGGCTGCATG 578
D 421 GCTGGAATCTTTGTAGAAACCATGAAAGAACTTTTCAAGCCTGGGAGGGCTGCATG 480
QY 579 AAAATTTCAGTCTGCTCTCCAGACACTTCGGCGGATCTTTTGGCCCTGTATGAGGTTG 638
D 481 AAAATTTCAGTCTGCTCTCCAGACACTTCGGCGGATCTTTTGGCCCTGTATGAGGTTG 540
QY 639 ATGACTTACGAGACGATTCGTACTTTGGGACTTTG 675
D 541 ATGACTTACGAGACGATTCGTACTTTGGGACTTTG 577
```

RESULT 41

```
AF529818
LOCUS Homo sapiens clone Ramos 4 truncated AID (AID) mRNA, complete cds.
DEFINITION AF529818
ACCESSION AF529818
VERSION AF529818.1 GI:22297223
KEYWORDS human.
```

ORGANISM

```
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 596)
Martin.A. and Scharff,M.D.
Somatic hypermutation of the AID transgene in B cells and non-B
cells
```

JOURNAL

```
Unpublished
2 (bases 1 to 596)
Martin,A. and Scharff,M.D.
Direct Submission
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
Location/Qualifiers
```

FEATURES

```
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 4"
1..596
/gene="AID"
1..570
/gene="AID"
/notes="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="truncated AID"
/protein_id="AAM95405.1"
/db_xref="GI:22297224"
/translocation="MDSLMNRKFLYQFNVRWAKRRRETYLCVYVKKRDSATSFSL
DFGLRNKNGCHVELFLRYISDWLDPCRCYRVTFWTSWSPCYDCARHVDFLRGNP
NLSRIPTARLYFCEDHKAPEGLRLHAGVQIALMFKDFYCNWTFVFNHETFK
AWEGLHNSVRLSRQLRLLPLYEVDLL"
```

gene

CDS

BASE COUNT

```
ORIGIN
128 a 163 c 154 g 151 t
Query Match
Best Local Similarity 18.3%; Score 516; DB 9; Length 596;
Matches 566; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCCTCTTGTATGAACCGGAGGAAGTTCTTTACCAATTTCAAAAATGTCGGTGG 139
D 1 ATGGACAGCCCTCTTGTATGAACCGGAGGAAGTTCTTTACCAATTTCAAAAATGTCGGTGG 60
QY 140 GCTAAGGTCGGCTGGAGACCTACTGTGTCTACGTAGTGAAGAGCGCTGACAGTGTCTACA 199
D 61 GCTAAGGTCGGCTGGAGACCTACTGTGTCTACGTAGTGAAGAGCGCTGACAGTGTCTACA 120
QY 200 TCCCTTTTCACTGGACTTTGGTTATCTTCCGAATAAGAACGCGCTGCCACGTGGAAATTCCTC 259
D 121 TCCCTTTTCACTGGACTTTGGTTATCTTCCGAATAAGAACGCGCTGCCACGTGGAAATTCCTC 180
QY 260 TTCTCTCCGTACATCTCGGACTTGGACCTAGACCCCTGGCCGCTGCTACCCGCTCACTGG 319
D 181 TTCTCTCCGTACATCTCGGACTTGGACCTAGACCCCTGGCCGCTGCTACCCGCTCACTGG 240
QY 320 TTCACTCTCTGGAGCCCTCTGTAGCACTGTGCCGACATGTGCCGACTTTCTGCGAGGG 379
D 241 TTCACTCTCTGGAGCCCTCTGTAGCACTGTGCCGACATGTGCCGACTTTCTGCGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAG 439
D 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360
QY 440 GCTGAGCCCGAGGGCTGCGCGCTGACCGCGCGGGTGCAGATAGTCAATAGTATCATGACC 499
D 361 GCTGAGCCCGAGGGCTGCGCGCTGACCGCGCGGGTGCAGATAGTCAATAGTATCATGACC 420
QY 500 TTCAAGATATTTTACTGCTGGAATACATTTTGTAGAAAACCATCAAAAGACTTTTCAAA 559
D 421 TTCAAGATATTTTACTGCTGGAATACATTTTGTAGAAAACCATCAAAAGACTTTTCAAA 480
QY 560 GCCTGGGAGGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGCGCATCTCTT 619
D 481 GCCTGGGAGGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGCGCATCTCTT 540
QY 620 TTCCCTCTGTATGAGGTTGATGACTTTA 646
D 541 TTCCCTCTGTATGAGGTTGATGACTTTA 567
```

RESULT 42

AF529817

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

```
AF529817
Homo sapiens clone Ramos 3 AID (AID) mRNA, partial cds.
AF529817.1 GI:22297221
```

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 596)

Martin,A. and Scharff,M.D.

Somatic hypermutation of the AID transgene in B cells and non-B

cells

Unpublished

2 (bases 1 to 596)

Martin,A. and Scharff,M.D.

Direct Submission

Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of

Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

Location/Qualifiers

1..596

/organism="Homo sapiens"

/db_xref="taxon:9606"

gene /clone="Ramos 3"
1..>596
/gene="AID"
CDS 1..>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95404.1"
/db_xref="GI:22297223"
/translation="MDSLMMNRKFLYOFKNVRWAKGPRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFYISDMDLDPGRYRVWTFWSNPSPCDCAHVAIDLGNP
NLSURIFTARLYFCEDRKAEPGRUHRAGVQIAIMTFKDDFYCWNTFVENHERTFK
AWEGHENSRLSRHLPLLYEVDLDAFRTLGL"
BASE COUNT 128 a 164 c 156 g 148 t

Query Match 17.5%; Score 494; DB 9; Length 596;
Best Local Similarity 99.7%; Pred. No. 3.9e-246;
Matches 594; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCGTGG 139
DB 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCGTGG 60
QY 140 GCTAAGGTCGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 199
DB 61 GCTAAGGTCGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 120
QY 200 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGAATTGCTC 259
DB 121 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGAATTGCTC 180
QY 260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGTACCGGTCACTGCG 319
DB 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGTACCGGTCACTGCG 240
QY 320 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCGCGGAGGG 379
DB 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCGCGGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAG 439
DB 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360
QY 440 GCTGAGCCCGAGGGCTGCGGCGGTGACCGCGCGGGGTGCAATAGCCATCATGACC 499
DB 361 GCTGAGCCCGAGGGCTGCGGCGGTGACCGCGCGGGGTGCAATAGCCATCATGACC 420
QY 500 TTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAAGTTCAAA 559
DB 421 TTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAAGTTCAAA 480
QY 620 TTGCCCTCTATGAGGTGATGACTTACGAGACGCAATTCGTAATTTGGGACTTTG 675
DB 541 TTGCCCTCTATGAGGTGATGACTTACGAGACGCAATTCGTAATTTGGGACTTTG 596

RESULT 43
AF529832
LOCUS
DEFINITION AF529832 clone 5 transgenic Homo sapiens AID (AID) mRNA, linear, 596 bp, mRNA, SYN 19-AUG-2002
ACCESSION AF529832
VERSION AF529832.1 GI:22297251
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 596)
Martin, A. and Scharff, M. D.
Somatic hypermutation of the AID transgene in B and non-B cells
Unpublished
2 (bases 1 to 596)
Martin, A. and Scharff, M. D.
Direct Submission
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Location/Qualifiers
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="5"
/cell_type="hybridoma PI-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95419.1"
/db_xref="GI:22297252"
/translation="MDSLMMNRKFLYOFKNVRWAKGPRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFYISDMDLDPGRYRVWTFWSNPSPCDCAHVAIDLGNP
NLSURIFTARLYFCEDRKAEPGRUHRAGVQIAIMTFKDDFYCWNTFVENHERTFK
AWEGHENSRLSRHLPLLYEVDLDAFRTLGL"
BASE COUNT 129 a 164 c 153 g 150 t
Query Match 17.5%; Score 494; DB 12; Length 596;
Best Local Similarity 99.7%; Pred. No. 3.9e-246;
Matches 594; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCGTGG 139
DB 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCGTGG 60
QY 140 GCTAAGGTCGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 199
DB 61 GCTAAGGTCGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 120
QY 200 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGAATTGCTC 259
DB 121 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGAATTGCTC 180
QY 260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGTACCGGTCACTGCG 319
DB 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGTACCGGTCACTGCG 240
QY 320 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCGCGGAGGG 379
DB 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCGCGGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAG 439
DB 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360
QY 440 GCTGAGCCCGAGGGCTGCGGCGGTGACCGCGCGGGGTGCAATAGCCATCATGACC 499
DB 361 GCTGAGCCCGAGGGCTGCGGCGGTGACCGCGCGGGGTGCAATAGCCATCATGACC 420
QY 500 TTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAAGTTCAAA 559
DB 421 TTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAAGTTCAAA 480
QY 560 GCCTGGGAGGCGTGCATGAAAATTCAGTTCTCTCCAGACAGCTTCGCGGCACTCTT 619
|||||

```
Db      481 GCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCACATCCTT 540
Qy      620 TTGCCCTCTATGAGCTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 675
Db      541 TTGCCCTCTATGAGCTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 596

RESULT 44
AF529836      596 bp      mRNA      linear      SYN 19-AUG-2002
LOCUS      Mus musculus clone 9 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION      complete cds.
ACCESSION      AF529836
VERSION      AF529836.1 GI:22297259
KEYWORDS      house mouse.
SOURCE      Mus musculus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS      1 (bases 1 to 596)
TITLE      Martin,A. and Scharff,M.D.
JOURNAL      Somatic hypermutation of the AID transgene in B and non-B cells
REFERENCE      2 (bases 1 to 596)
AUTHORS      Martin,A. and Scharff,M.D.
TITLE      Direct Submission
JOURNAL      Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES      Location/Qualifiers
source      1..596
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /clone="9"
              /cell_type="hybridoma P1-5"
              /transgenic
source      1..596
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
gene      1..>596
              /gene="AID"
CDS      1..>596
              /gene="AID"
              /codon_start=1
              /product="AID"
              /protein_id="AAM95423.1"
              /db_xref="GI:22297260"
              /translation="MDSLMMNRKFLYQFKNVRWSKGRRETYLCYVVKRRDSATSFSL
              DGYLRNNGCHVELLFRLYISDWLDPCRCYRVHTFTSWSPCYDCARHVADFLRGNP
              NLSLRFTARLYFCEDRKAEPGLRLRHAGVOIAIMTKDYFYCWNTFVENHERTEK
              AWEGLHENSVRLSRQLRLLPLLYEVDLDRDAFTLGL"
BASE COUNT      127 a 164 c 153 g 152 t
ORIGIN

Query Match      17.5%; Score 494; DB 12; Length 596;
Best Local Similarity      99.7%; Pred. No. 3.9e-246;
Matches 594; Conservative      0; Mismatches      2; Indels      0; Gaps      0;

Qy      80 ATGNACAGCCCTCTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 139
Db      1 ATGGACAGCCCTCTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60

Qy      140 GCTAAGGTCGGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGCTGACAGTGCTACA 199
Db      61 TCTAAGGTCGGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGCTGACAGTGCTACA 120

Qy      200 TCCCTTTTACATGGACTTTGGTTATCTTCGCAATAAGAACGCGCTGCCACGTGGAAATTCGTC 259
Db      121 TCCCTTTTACATGGACTTTGGTTATCTTCGCAATAATAACGCGCTGCCACGTGGAAATTCGTC 180

Qy      260 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 319
Db      181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 240
```

```
Qy      320 TTCACTCTCTGAGAGCCCTGCTAGACTGTGCGCAGACATGTGGCGGACTTTTCTGCGAGG 379
Db      241 TTCACTCTCTGAGAGCCCTGCTAGACTGTGCGCAGACATGTGGCGGACTTTTCTGCGAGG 300
Qy      380 AACCCCAACCTCAGTCTCAGGATCTTACCGCGCGCCTCTACTTCTGTGAGGACCGCAAG 439
Db      301 AACCCCAACCTCAGTCTCAGGATCTTACCGCGCGCCTCTACTTCTGTGAGGACCGCAAG 360
Qy      440 GCTGAGCCGAGGGGCTGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 499
Db      361 GCTGAGCCGAGGGGCTGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
Qy      500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAAA 559
Db      421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAAA 480
Qy      560 GCCTGGGAAGCGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGCGCATCTTT 619
Db      481 GCCTGGGAAGCGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGCGCATCTTT 540
Qy      620 TTGCCCTCTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 675
Db      541 TTGCCCTCTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG 596

RESULT 45
AF529838      596 bp      mRNA      linear      SYN 19-AUG-2002
LOCUS      Mus musculus clone 11 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION      complete cds.
ACCESSION      AF529838
VERSION      AF529838.1 GI:22297263
KEYWORDS      house mouse.
SOURCE      Mus musculus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS      1 (bases 1 to 596)
TITLE      Martin,A. and Scharff,M.D.
JOURNAL      Somatic hypermutation of the AID transgene in B and non-H cells
REFERENCE      2 (bases 1 to 596)
AUTHORS      Martin,A. and Scharff,M.D.
TITLE      Direct Submission
JOURNAL      Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES      Location/Qualifiers
source      1..596
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /clone="11"
              /cell_type="hybridoma P1-5"
              /transgenic
source      1..596
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
gene      1..>596
              /gene="AID"
CDS      1..>596
              /gene="AID"
              /codon_start=1
              /product="AID"
              /protein_id="AAM95425.1"
              /db_xref="GI:22297264"
              /translation="MDSLMMNRKFLYQFKNVRWAKRRRETYLCYVVKRRDSATSFSL
              DEGYLRNNGCHVELLFRLYISDWLDPCRCYRVHTFTSWSPCYDCARHVADFLRGNP
              NLSLRFTARLYFCEDRKAEPGLRLRHAGVOIATMTKDYFYCWNTFVENIERTEK
              AWEGLHENSVRLSRQLRLLPLLYEVDLDRDAFTLGL"
BASE COUNT      128 a 163 c 154 g 151 t
ORIGIN

Query Match      17.5%; Score 494; DB 12; Length 596;
Best Local Similarity      99.7%; Pred. No. 3.9e-246;
```

```
Matches 594; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCTGG 60
QY 140 GCTAAGGGTGGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 199
Db 61 GCTAAGGGTGGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 120
QY 200 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATGTCTC 259
Db 121 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATGTCTC 180
QY 260 TTCCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCGTGTCTACCGCGCTCACCTGG 319
Db 181 TTCCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCGTGTCTACCGCGCTCACCTGG 240
QY 320 TTCACCTCTCTGGAGCCCTGTGCTACGACTGTGCCCGACATGTGGCCGACTTCTTGGAGGG 379
Db 241 TTCACCTCTCTGGAGCCCTGTGCTACGACTGTGCCCGACATGTGGCCGACTTCTTGGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCCGAGGGGTGGCGGGTGCACCGCGCGGCTGCAATAGCCATCATGACC 499
Db 361 GCTGAGCCCGAGGGGTGGCGGGTGCACCGCGCGGCTGCAATAGCCATCATGACC 420
QY 500 TTCAAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTTTCAA 559
Db 421 TTCAAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTTTCAA 480
QY 560 GCCTGGAAGGGCTGCATGAAATTCAGTTTCTCTCTCAGACAGCTTCGGCGCATCCTT 619
Db 481 GCCTGGAAGGGCTGCATGAAATTCAGTTTCTCTCTCAGACAGCTTCGGCGCATCCTT 540
QY 620 TTGCCCTCTGTAGAGTTTGATGACTTACGAGAGCGCATTTCTGACTTTGGGACTTTG 675
Db 541 TTGCCCTCTGTAGAGTTTGATGACTTACGAGAGCGCATTTCTGACTTTGGGACTTTG 596

RESULT 46
AF529856 547 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 16 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529856
VERSION AF529856.1 GI:22297299
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source Location/Qualifiers
1. .547
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="16"
/cell_line="CHO"
/transgenic
```

```
1. .547
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1. .547
/gene="AID"
<1. .547
/gene="AID"
/codon_start=2
/product="AID"
/protein_id="AAM5443.1"
/db_xref="GI:22297300"
/translation="VRWAKRRRETYLCYVVKRRDSATSFSLDFGLRNKNKGHVLLF
LRYISDWLDPRCYRYVTFWTFWSPCYCARHVDLFRGNPNLSLRIFTARLYFCEDR
KAEPEGLRLRHRAQVQIAIMTFKDYFCYWNTEFVENHERTEKAWEGLEHNSVRLSRQLR
RIILPLYEVDDLRLDAFTWGR"
BASE COUNT 112 a 153 c 147 g 135 t
ORIGIN
Query Match 17.3%; Score 487; DB 12; Length 547;
Best Local Similarity 100.0%; Pred. No. 1.7e-242;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 130 TGTCCGCTGGGCTAAGGCTCGCGGTGAGACCTACCTGTGCTAGTCAAGAGCGTGA 189
Db 1 TGTCCGCTGGGCTAAGGCTCGCGGTGAGACCTACCTGTGCTAGTCAAGAGCGTGA 60
QY 190 CAGTGCTACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGT 249
Db 61 CAGTGCTACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGT 120
QY 250 GGAATTTGCTTCTTCGCGTACATCTCGGACTGGGACCTAGACCTGGCCGCTGTCTACCG 309
Db 121 GGAATTTGCTTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGTCTACCG 180
QY 310 CGTCACTCGGTTCACTCCTCTGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTT 369
Db 181 CGTCACTCGGTTCACTCCTCTGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTT 240
QY 370 TCTGGGAGGAAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGA 429
Db 241 TCTGGGAGGAAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGA 300
QY 430 GGACCGCAAGGCTGAGCCCGAGGGGCTGGCGGCTGACCCGCGGGGTGCAATAGC 489
Db 301 GGACCGCAAGGCTGAGCCCGAGGGGCTGGCGGCTGACCCGCGGGGTGCAATAGC 360
QY 490 CATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAG 549
Db 361 CATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAG 420
QY 550 AACTTTTCAAAGCTGGGAAGGCTGCATGAAAATTCAGTTCTCTCTCCAGACAGCTTCG 609
Db 421 AACTTTTCAAAGCTGGGAAGGCTGCATGAAAATTCAGTTCTCTCTCCAGACAGCTTCG 480
QY 610 GCGCATC 616
Db 481 GCGCATC 487

RESULT 47
AC007240/c
LOCUS Homo sapiens BAC clone RP11-83M8 from 2, complete sequence.
DEFINITION AC007240
ACCESSION AC007240.2 GI:5306303
VERSION
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
```


repeat_region /rpt_family="Alu" 17799. .18095
repeat_region /rpt_family="Alu" 18146. .18449
repeat_region /rpt_family="L1" 18450. .18745
repeat_region /rpt_family="Alu" 18746. .18928
repeat_region /rpt_family="L1" 18929. .19230
repeat_region /rpt_family="Alu" 19231. .19570
repeat_region /rpt_family="L1" 19571. .19880
repeat_region /rpt_family="Alu" 19881. .20109
repeat_region /rpt_family="L1" 20110. .20415
repeat_region /rpt_family="Alu" 20416. .20446
repeat_region /rpt_family="L1" 20446. .20677
repeat_region /rpt_family="L1" 20683. .20974
repeat_region /rpt_family="Alu" 21002. .21107
repeat_region /rpt_family="L1" 21464. .21698
repeat_region /rpt_family="L1" 21825. .22150
repeat_region /rpt_family="Alu" 22322. .22626
repeat_region /rpt_family="Alu" 23770. .23972
repeat_region /rpt_family="MER1_type" 24272. .24294
repeat_region /rpt_family="CAAT)n" 24304. .24576
repeat_region /rpt_family="Alu" 24646. .24962
repeat_region /rpt_family="Alu" 24973. .25410

Query Match 3.1%; Score 86; DB 9; Length 203250;

Best Local Similarity 100.0%; Pred. No. 2.le-33;

Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1970 GCGGTGGTACGAGGACCTGTAATCCAGCTACTTGGGAGGCTGAGCGAGGAGTGCCT 2029
DB 18606 GCGGTGGTACGAGGACCTGTAATCCAGCTACTTGGGAGGCTGAGCGAGGAGTGCCT 18547
QY 2030 TCAACCCAGGAGGTGGAGTTGCAGT 2055
DB 18546 TGAACCCAGGAGGTGGAGTTGCAGT 18521

RESULT 48
AC079329 LOCUS 156540 bp DNA linear PRI 20-APR-2002
DEFINITION Homo sapiens chromosome 11, clone RP11-265D17, complete sequence.
ACCESSION AC079329
VERSION AC079329.15 GI:20128016
KEYWORDS HTC.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 156540)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-265D17
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 156540)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L.,
Boukhaltier,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Lacroque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G.,
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo.A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Submitted (27-AUG-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 156540)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhaltier,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collymore,A.,
Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Jones,C.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Olivier,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 156540)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhaltier,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collymore,A.,
Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Olivier,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL

Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome

COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 10, 2002 this sequence version replaced gi:19683125.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L10912
 Center clone name: 265_D_17

FEATURES

source

Location/Qualifiers

1..156540
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11"
 /clone="RP11-265D17"
 /clone_lib="RPC1-11 Human Male BAC"
 134..139
 /note="<30 qual SINGL region"
 813..879
 /note="single clone coverage"
 1951..2104
 /rpt_family="MIR"
 complement(2247..2542)
 /rpt_family="AluX"
 complement(2778..2877)
 /rpt_family="L3"
 complement(2885..3564)
 /rpt_family="MER104A"
 complement(3566..4179)
 /rpt_family="L3"
 complement(4218..4323)
 /rpt_family="L3"
 complement(5098..5386)
 /rpt_family="AluSg"
 complement(6406..6481)
 /rpt_family="L2"
 6539..6559
 /rpt_family="AT_Rich"
 complement(7143..7430)
 /rpt_family="L2"
 7708..7854
 /rpt_family="MIR"
 complement(8134..8319)
 /rpt_family="MIR"
 8584..8784
 /rpt_family="MIR"
 8798..8826
 /rpt_family="AT_Rich"
 complement(9140..9272)
 /rpt_family="MIR"
 9455..9489
 /rpt_family="(T)n"
 complement(9549..9844)
 /rpt_family="AluJo"
 10846..11158
 /rpt_family="AluYb8"
 11570..11783
 /rpt_family="MIR"
 complement(11746..11929)
 /rpt_family="MIR"
 complement(12908..13215)
 /rpt_family="AluX"
 13878..14048
 /rpt_family="MIR"
 complement(14481..14549)
 /rpt_family="MLT1K"
 14554..14673

repeat_region /rpt_family="AluJo"
 complement(15075..15285)
 /rpt_family="MIR"
 repeat_region complement(15300..15549)
 /rpt_family="MIR"
 repeat_region complement(15793..15857)
 /rpt_family="L2"
 16085..16320
 /rpt_family="MIR"
 complement(16384..16463)
 /rpt_family="MIR3"
 17431..17457
 /rpt_family="(TG)n"
 18440..18652
 /rpt_family="L2"
 complement(18894..19027)
 /rpt_family="MIR3"
 repeat_region complement(20058..20252)
 /rpt_family="MIR3"
 20274..20323
 /rpt_family="L2"
 repeat_region 21013..21198
 /rpt_family="MIR3"
 repeat_region complement(21757..21948)
 /rpt_family="MIR3"
 repeat_region complement(22497..22689)
 /rpt_family="MIR"
 repeat_region complement(22767..23110)
 /rpt_family="L2"
 23765..23795
 /rpt_family="(GA)n"
 25844..27056
 /rpt_family="L1PA5"
 repeat_region complement(27157..27230)
 /rpt_family="L2"
 27536..27612
 /rpt_family="L2"
 repeat_region complement(27626..27783)
 /rpt_family="MERSB"

Query Match 2.8% Score 79; DB 9; Length 156540;

Best Local Similarity 100.0%; Pred. No. 9.2e-30;

Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1981 AGGCACCTGTAATCCAGCTACTTGGGAGGCTGAGGAGAGAGATCGCTTCAAGCCAGGA 2040
 Db 32001 AGGCACCTGTAATCCAGCTACTTGGGAGGCTGAGGAGAGAGATCGCTTCAAGCCAGGA 32060
 Qy 2041 GGTGGAGGTTGCAGTAAGC 2059
 Db 32061 GGTGGAGGTTGCAGTAAGC 32079

RESULT 49

AC023634/c

LOCUS

DEFINITION Homo sapiens chromosome 11 clone RP11-133112 map 11, WORKING DRAFT
 AC023634 187836 bp DNA linear HTG 16-MAR-2001
 SEQUENCE, 8 unordered pieces.

AC023634

AC023634.3 GI:13357436

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS Homo sapiens.

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 187836)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens chromosome 11, clone RP11-133112

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 187836)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Bada, F., Boquslavsky, I.,

Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castilo, A.,

ChoepeI,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,K., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,K., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (16-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2001 this sequence version replaced gi:7767803.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WTBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence-submissions@genome.wi.mit.edu

----- Project Information

Center project name: I5395

Center clone name: I33.I.12

----- Summary Statistics

Sequencing vector: M13; M77815; 56% of reads

Sequencing vector: Plasmid; n/a; 44% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 179213 bases at least Q40

Consensus quality: 181746 bases at least Q30

Consensus quality: 184176 bases at least Q20

Insert size: 174000; agarose-fp

Insert size: 187136; sum-of-contigs

Quality coverage: 13.0 in Q20 bases; agarose-fp

Quality coverage: 12.1 in Q20.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 8 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 33786: contig of 33786 bp in length

* 33787 33886: gap of 100 bp

* 33887 35099: contig of 1213 bp in length

* 35100 35199: gap of 100 bp

* 35200 36514: contig of 1315 bp in length

* 36515 36614: gap of 100 bp

* 36615 37990: contig of 1376 bp in length

* 37991 38090: gap of 100 bp

* 38091 39599: contig of 1509 bp in length

* 39600 39699: gap of 100 bp

* 39700 41135: contig of 1436 bp in length

* 41136 41235: gap of 100 bp

* 41236 43376: contig of 2141 bp in length

* 43377 43476: gap of 100 bp

* 43477 187836: contig of 144360 bp in length.

Location/Qualifiers

1..187836

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="11"

FEATURES
source

/map="11"
/clone="RP11-133112"
/clone_lib="RP11 Human Male BAC"
1..33786
/note="assembly_fragment"
vector_side:left
33887..35099
/note="assembly_fragment"
35200..36514
/note="assembly_fragment"
36615..37990
/note="assembly_fragment"
38091..39599
/note="assembly_fragment"
39700..41135
/note="assembly_fragment"
41236..43376
/note="assembly_fragment"
43477..187836
/note="assembly_fragment"
clone_end:T7
vector_side:right
BASE COUNT 51537 a 44861 c 42761 g 47977 t 700 others
ORIGIN
Query Match 2.8%; Score 79; DB 2; Length 187836;
Best Local Similarity 100.0%; Pred No. 9.2e-30;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1981 AGGCACCTGTAATCCAGCTTCTGGAGGCTGAGCAGGAGTAATCGTTGAACCCAGGA 2040
|||||
Db 76398 AGGCACCTGTAATCCAGCTTCTGGAGGCTGAGCAGGAGTAATCGTTGAACCCAGGA 76339
QY 2041 GGTGGAGGTTGCAGTAAGC 2059
|||||
Db 76338 GGTGGAGGTTGCAGTAAGC 76320
AC004603
LOCUS AC004603 41511 bp DNA linear PRI 23-APR-1998
DEFINITION Homo sapiens chromosome 19, cosmid R33496, complete sequence.
ACCESSION AC004603
VERSION AC004603.1 GI:3077822
KEYWORDS HTG.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 41511)
Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Adamson,A.W.,
Burkhardt-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,
Phan,H., Velasco,N., Ganes,J., Danganan,L., Poundstone,P.,
Christensen,M., Georgescu,A., Avila,J., Liu,S., Attix,C.,
Andreise,T., Frankheim,M., Anico-Keller,G., Coefield,J., Duarte,S.,
Lucas,S., Bruce,R., Thomas,P., Quan,G., Kronmiller,B., Arellano,A.,
Montgomery,M., Ow,D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.O.
and Carrano,A.V.
TITLE Sequence analysis of a 2.5 Mb region in 19ql3.2 containing a
clustered CEA/PSG gene family
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 41511)
Lamerdin,J.E.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT Map and sequence oriented from q centromere to telomere. Cosmid
R33496 overlaps cosmid F19987 to the right (by 2.447 kb) and cosmid
F25173 to the right. Additional chr 19 map and sequence information
may be obtained at:
<http://www-bio.llnl.gov/bbrp/genome/genome.html>.
Location/Qualifiers

FEATURES

```
source
1. .41511
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.2 from BCKDHA-D19S217"
/clone="R33496"
/cell_line="5H12-B"
/clone_lib="LL19NC03 R chromosome 19-specific cosmid library"
/notes="Cosmid library constructed at LLNL from flow-sorted chromosomes from hybrid 5H12-B, which carries chromosome 19 as its only human chromosome."
complement(1. .616)
/rpt_family="L1"
612. .938
/rpt_family="L1"
934. .2630
/rpt_family="L1"
3217. .3631
/rpt_family="MSTB"
complement(5887. .5918)
/rpt_family="AT-rich"
complement(6303. .6343)
/rpt_family="(TAAA)n"
complement(6365. .6650)
/rpt_family="AluJo"
6793. .7179
/rpt_family="MLT1B"
9090. .9194
/rpt_family="LINE2"
complement(9305. .9408)
/rpt_family="FLAW_A"
9878. .10178
/rpt_family="AluYa5"
complement(10336. .10493)
/rpt_family="L1ME"
10676. .10792
/rpt_family="LINE2"
10954. .11356
/rpt_family="MSTB"
11886. .12076
/notes="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: good, score: 65.000"
/rpt_family="MIR"
12353. .12393
12727. .12945
/rpt_family="L1ME"
12969. .13742
/rpt_family="L1MB3"
13743. .14041
/rpt_family="AluY"
14061. .14108
/rpt_family="L1MB4"
14567. .14361
/notes="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: good, score: 62.000"
/rpt_family="MIR"
complement(15609. .16217)
/rpt_family="MER67C"
16221. .16426
/rpt_family="AluJo"
16427. .16457
/rpt_family="(TGAA)n"
complement(16458. .16509)
/rpt_family="MER67C"
16548. .16661
/notes="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: good, score: 72.000"
/rpt_family="AluSx"
17689. .17990
18091. .18371
/rpt_family="HERVL"
18445. .18540
/rpt_family="MLT1B"

repeat_region complement(18583. .19157)
/rpt_family="L1"
repeat_region complement(19158. .19458)
/rpt_family="AluY"
19890. .19924
/rpt_family="(CA)n"
repeat_region complement(19925. .20055)
/rpt_family="MER4_internal"
20076. .20439
/rpt_family="MER50"
repeat_region complement(20469. .20524)
/rpt_family="(TA)n"
20541. .20823
/rpt_family="MER4_internal"
complement(20869. .21760)
/rpt_family="L1PA8"
22250. .22542
/rpt_family="AluJo"
22812. .23005
/rpt_family="MER4C"
23594. .23733
/notes="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 83.000"
23753. .23899
/rpt_family="MIR"
complement(24669. .24849)
/rpt_family="MIR"
24950. .24985
/rpt_family="MLT1D"
24987. .25287
/rpt_family="AluSg"
25291. .25469
/rpt_family="MLT1D"
complement(25608. .25638)
/rpt_family="AluSg"
complement(25856. .26042)
/rpt_family="MER11A"
complement(26223. .26594)
/rpt_family="HERVK"
complement(26739. .26902)
/rpt_family="AluSg"
complement(26911. .27277)
/rpt_family="MER57_internal"
27440. .27732
/rpt_family="AluSp"
complement(27759. .29229)
/rpt_family="MER65_internal"
complement(29238. .29530)
/rpt_family="AluY"
complement(29531. .29623)
/rpt_family="MER65_internal"
29819. .29969
/rpt_family="MER5B"
complement(30951. .30994)
/notes="Exon, pregnancy-specific beta-1-glycoprotein 5-like pseudogene. DPS similarity to M32635) pregnancy-specific beta-1-glycoprotein 5 [Homo sapiens] (residues 500. .512); 67% identity."
/pseudo
31507. .31749
/rpt_family="LTR16B"
complement(31877. .32061)
/rpt_family="MIR"
complement(32142. .32243)
/rpt_family="MIR"
32244. .32543
/rpt_family="L1"
32546. .32860
/rpt_family="AluJo"
32862. .32986
/rpt_family="L1"
33029. .35071
/rpt_family="L1MB4"
```

```

exon      complement(35975. .36229)
/note="Exon, pregnancy-specific beta-1-glycoprotein 5-like
pseudogene. Dps similarity to (M32635) pregnancy-specific
beta-1-glycoprotein 5 [Homo sapiens] (residues 415. .499);
91% identity."
/pseudo
exon      complement(36632. .36910)
/note="Exon, pregnancy-specific beta-1-glycoprotein 5-like
pseudogene. Dps similarity to (M32635) pregnancy-specific
beta-1-glycoprotein 5 [Homo sapiens] (residues 322. .414);
91% identity."
/pseudo
repeat_region complement(37853. .37911)
/rpt_family="LIMB1"
repeat_region 37924. .38055
/rpt_family="LIMB1"
repeat_region 38056. .38218
/rpt_family="MER46"
repeat_region complement(38485. .38798)
/rpt_family="L1"
exon      complement(38984. .39238)
/note="Exon, pregnancy-specific beta-1-glycoprotein 5-like

Query Match      2.7%; Score 77; DB 9; Length 41511;
Best Local Similarity 100.0%; Pred. No. 1e-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGGCACCTGTATCCCGACTTGGGAGGCTGAGGAGGAGGAGTCCGTTGAACCCAG 2038
|||||
Db 1788 GCAGGCACCTGTATCCCGACTTGGGAGGCTGAGGAGGAGGAGTCCGTTGAACCCAG 17897
|||||

QY 2039 GAGGTGGAGGTTGCAGT 2055
|||||
Db 17898 GAGGTGGAGGTTGCAGT 17914
|||||

RESULT 51
AC084034/c
LOCUS
DEFINITION
AC084034 103287 bp DNA linear HTG 09-MAY-2002
HOMO sapiens chromosome 3 clone RP11-502H22, *** SEQUENCING IN
PROGRESS ***, 30 unordered pieces.
AC084034
AC084034.15 GI:20335777
HTG: HTGS_PHASE1.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J.J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carrolli,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karissom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

```

```

TITLE
JOURNAL
REFERENCE 2 (bases 1 to 103287)
AUTHORS Worley,K.C.
TITLE
JOURNAL
REFERENCE 3 (bases 1 to 103287)
AUTHORS Worley,K.C.
TITLE
JOURNAL
COMMENT
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCGG
Center clone name: RP11-502H22
----- Summary Statistics
Sequencing vector: M13;
Chemistry: Dye-primer Bodipy: 37% of reads
Chemistry: Dye-terminator Big Dye: 63% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 117633 bases at least Q40
Consensus quality: 136061 bases at least Q30
Consensus quality: 148159 bases at least Q20
Estimated insert size: 141518; sum-of-contigs estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2194: contig of 2194 bp in length
* 2195 2294: gap of unknown length
* 2295 4434: contig of 2140 bp in length
* 4435 4534: gap of unknown length
* 4535 6710: contig of 2176 bp in length
* 6711 6810: gap of unknown length
* 6811 9027: contig of 2217 bp in length
* 9028 9127: gap of unknown length
* 9128 11310: contig of 2183 bp in length
* 11311 11410: gap of unknown length
* 11411 13930: contig of 2520 bp in length
* 13931 14030: gap of unknown length
* 14031 16531: contig of 2501 bp in length

```


repeat_region 282. .668 /rpt_family="MaLR"
repeat_region 1575. .1615 /rpt_family="AT_rich"
repeat_region 1616. .1897 /rpt_family="Alu"
repeat_region 3392. .3775 /rpt_family="Retroviral"
misc_feature 4319. .4868 /note="match to EST AI568539 (NID:g4531913) tn40e08.xl1"
repeat_region 5673. .5921 /rpt_family="MER1_type"
misc_feature 8043. .10028 /note="CpG_island (%GC=68.4, o/e=0.90, #CpGs=214)"
repeat_region 8332. .8364 /rpt_family="A-rich"
repeat_region 9202. .9330 /rpt_family="CCCCG)n"
repeat_region 9950. .9970 /rpt_family="GC_rich"
repeat_region 10090. .10212 /rpt_family="L1"
repeat_region 10986. .11122 /rpt_family="(TAGA)n"
repeat_region 11243. .11360 /rpt_family="(CAGA)n"
repeat_region 12015. .12240 /rpt_family="MER4-group"
repeat_region 12243. .12853 /rpt_family="MER4-group"
repeat_region 12881. .12991 /rpt_family="MER94"
repeat_region 12998. .13524 /rpt_family="L1"
repeat_region 16196. .16425 /rpt_family="Retroviral"
repeat_region 17256. .17613 /rpt_family="Retroviral"
repeat_region 18007. .18149 /rpt_family="MER2_type"
repeat_region 18224. .18522 /rpt_family="MER2_type"
repeat_region 18715. .18832 /rpt_family="L1"
repeat_region 19286. .19373 /rpt_family="L2"
repeat_region 20391. .20481 /rpt_family="MIR"
repeat_region 21808. .22111 /rpt_family="Alu"
repeat_region 22147. .22188 /rpt_family="AT_rich"
repeat_region 22392. .22555 /rpt_family="MaLR"
repeat_region 23127. .23212 /rpt_family="MaLR"
repeat_region 24894. .25083 /rpt_family="MER4-group"
repeat_region 25147. .25237 /rpt_family="MER4-group"
repeat_region 25269. .25598 /rpt_family="MER4-group"
repeat_region 25616. .26607 /rpt_family="MER4-group"
repeat_region 26945. .27063 /rpt_family="MER4-group"
repeat_region 27210. .27502 /rpt_family="Alu"
repeat_region 27503. .27768 /rpt_family="MER4-group"
repeat_region 28822. .29089 /rpt_family="MER4-group"
repeat_region 29111. .29368

repeat_region 29377. .29464 /rpt_family="MER4-group?"
repeat_region 29748. .29786 /rpt_family="AT_rich"
repeat_region 29893. .30466 /rpt_family="L1"
repeat_region 30524. .30638 /rpt_family="(CATA)n"
repeat_region 30886. .30997 /rpt_family="(CA)n"
repeat_region 31015. .31125 /rpt_family="(CAATA)n"
repeat_region 31099. .31230 /rpt_family="(CACAC)n"
repeat_region 32285. .33146 /rpt_family="L1"
repeat_region 33240. .33373 /rpt_family="MER2_type"
repeat_region 33380. .33559 /rpt_family="L1"
misc_feature 35690. .37077 /note="CpG_island (%GC=63.0, o/e=0.90, #CpGs=106)"
repeat_region 37890. .40107 /rpt_family="MER2_type"
repeat_region 42804. .43109 /rpt_family="Alu"
repeat_region 44304. .44361 /rpt_family="MIR"

Query Match 2.7%; Score 77; DB 9; Length 112804;
Best Local Similarity 100.0%; Pred.No.le=28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGGCACCTGTAATCCAGCTACTTGGAGGCTGAGCAGGAGTAATCGCTTGAACCCAG 2038
|||||
Db 101773 GCAGGCACCTGTAATCCAGCTACTTGGAGGCTGAGCAGGAGTAATCGCTTGAACCCAG 101714
|||||

QY 2039 GAGGTGGAGTTGCAGT 2055
|||||
Db 101713 GAGGTGGAGTTGCAGT 101697
|||||

RESULT 55
LOCUS AY052369 121028 bp DNA linear PRI 16-APR-2002
DEFINITION Homo sapiens PP2A B56 gamma gene, complete cds, alternative splice products.
ACCESSION AY052369
VERSION AY052369.1 GI:16303629
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 121028)
AUTHORS Muneer,S., Ramalingam,V., Wyatt,R., Schultz,R.A., Minna,J.D. and Kamibayashi,C.
TITLE Genomic organization and mapping of the gene encoding the PP2A B56gamma regulatory subunit
JOURNAL Genomics 79 (3), 344-348 (2002)
MEDLINE 21853305
PUBMED 11863364
REFERENCE 2 (bases 1 to 121028)
AUTHORS Muneer,S., Kamibayashi,C. and Minna,J.D.
TITLE Direct Submission
JOURNAL Submitted (24-AUG-2001) Hamon Center for Therapeutic Oncology, University of Texas Southwestern Medical Center, 6000 Harry Hines Blvd, Dallas, TX 75390, USA
FEATURES
Location/Qualifiers
1..121028
/db_xref="taxon:9606"

```

/chromosome="14"
/map="14q32.2"
join(2507..2643,49292..49491,74767..74877,75863..75955,
76038..76168,82840..82899,85603..85711,87110..87163,
94322..94492,99005..99132,102192..102293,105004..105076,
110434..110550,117745..120289)
/product="PP2A B56 gamma 3"
join(2507..2643,49292..49491,74767..74877,75863..75955,
76038..76168,82840..82899,85603..85711,87110..87163,
94322..94492,99005..99132,102192..102293,105004..105076,
117745..120289)
/product="PP2A B56 gamma 2"
join(2507..2643,49292..49491,74767..74877,75863..75955,
76038..76168,82840..82899,85603..85711,87110..87163,
94322..94492,99005..99132,102192..102293,105004..105226)
/product="PP2A B56 gamma 1"
2507..2643
/number=1
join(2580..2643,49292..49491,74767..74877,75863..75955,
76038..76168,82840..82899,85603..85711,87110..87163,
94322..94492,99005..99132,102192..102293,105004..105076,
110434..110550,117745..117876)
/feature="protein phosphatase; splice variant 3"
/codon_start=1
/product="PP2A B56 gamma 3"
/protein_id="AAL14778.1"
/db_xref="GI:16303632"
/translation="MVDAANSNGFPQPVLLHIRDVPPADQEKLFIOKLQCCVLFD
FVSDPLSKKVEKRAALSEMVEYITHNRNVITEPIYEVVHMFVAVNFTLPPSSN
PTGAFDPEDEPTLEAAPHQLQVYEFLEFLESDFQPNIAKKYIDQKVLQLEL
FDSDFPRDLTKTLHRIYKGLRAYIRKQINNIYRIYETEHHNGIAELLEIL
GSIINGFALPKKEHKIFLLKVLPLHKVKSLSVYHQLAYCVQVLEKDSLTLEPVV
MALLYWPKTHSPKEVMFLNELEILDIPESEFVKIMEPLFROLAKCVSSPHFOVAE
RALLYWNNYETMSLISDAAKILPIMFPSLYRNSKTHNKTIHGLIYNALKLFEMNQ
KLFDCTQOQFAEKLEKLMKEREAWKIENTAKANPOYTVSOASTMSIPVAMET
DGFLFEDVQMLRKYKDBAQKDPKRAKDRPLALRKSLEPQDPHFKKALEAHCRADEL
ASQDGR"
join(2580..2643,49292..49491,74767..74877,75863..75955,
76038..76168,82840..82899,85603..85711,87110..87163,
94322..94492,99005..99132,102192..102293,105004..105076,
117745..117876)
/feature="protein phosphatase; splice variant 2"
/codon_start=1
/product="PP2A B56 gamma 2"
/protein_id="AAL14778.1"
/db_xref="GI:16303631"
/translation="MVDAANSNGFPQPVLLHIRDVPPADQEKLFIOKLQCCVLFD
FVSDPLSKKVEKRAALSEMVEYITHNRNVITEPIYEVVHMFVAVNFTLPPSSN
PTGAFDPEDEPTLEAAPHQLQVYEFLEFLESDFQPNIAKKYIDQKVLQLEL
FDSDFPRDLTKTLHRIYKGLRAYIRKQINNIYRIYETEHHNGIAELLEIL
GSIINGFALPKKEHKIFLLKVLPLHKVKSLSVYHQLAYCVQVLEKDSLTLEPVV
MALLYWPKTHSPKEVMFLNELEILDIPESEFVKIMEPLFROLAKCVSSPHFOVAE
RALLYWNNYETMSLISDAAKILPIMFPSLYRNSKTHNKTIHGLIYNALKLFEMNQ
KLFDCTQOQFAEKLEKLMKEREAWKIENTAKANPOYTVSOASTMSIPVAMET
DGFLFEDVQMLRKYKDBAQKDPKRAKDRPLALRKSLEPQDPHFKKALEAHCRADEL
ASQDGR"
join(2580..2643,49292..49491,74767..74877,75863..75955,
76038..76168,82840..82899,85603..85711,87110..87163,
94322..94492,99005..99132,102192..102293,105004..105100)
/feature="protein phosphatase; splice variant 1"
/codon_start=1
/product="PP2A B56 gamma 1"
/protein_id="AAL14777.1"
/db_xref="GI:16303630"
/translation="MVDAANSNGFPQPVLLHIRDVPPADQEKLFIOKLQCCVLFD
FVSDPLSKKVEKRAALSEMVEYITHNRNVITEPIYEVVHMFVAVNFTLPPSSN
PTGAFDPEDEPTLEAAPHQLQVYEFLEFLESDFQPNIAKKYIDQKVLQLEL
FDSDFPRDLTKTLHRIYKGLRAYIRKQINNIYRIYETEHHNGIAELLEIL
GSIINGFALPKKEHKIFLLKVLPLHKVKSLSVYHQLAYCVQVLEKDSLTLEPVV
MALLYWPKTHSPKEVMFLNELEILDIPESEFVKIMEPLFROLAKCVSSPHFOVAE
RALLYWNNYETMSLISDAAKILPIMFPSLYRNSKTHNKTIHGLIYNALKLFEMNQ
KLFDCTQOQFAEKLEKLMKEREAWKIENTAKANPOYTVSOASTMSIPVAMET
DGFLFEDVQMLRKYKDBAQKDPKRAKDRPLALRKSLEPQDPHFKKALEAHCRADEL
ASQDGR"
46492..49491
/number=2

```

```

exon
74767..74877
/number=3
exon
75863..75955
/number=4
exon
76038..76168
/number=5
exon
82840..82899
/number=6
exon
85603..85711
/number=7
exon
87110..87163
/number=8
exon
94322..94492
/number=9
exon
99005..99132
/number=10
exon
102192..102293
/number=11
exon
105004..105076
/number=12
exon
105077..105226
/number=12A
exon
110434..110550
/number=13
exon
117745..120289
/number=14
BASE COUNT 31520 a 26430 c 27189 g 35887 t 2 others
ORIGIN
Query Match 2.7% Score 77: DB 9: Length 121028:
Best Local Similarity 100.0% Pred. No. le-28;
Matches 77: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
QY 1979 CGAGCACCCTGTAAATCCAGCTACTTTGGAGGCTCAGGACGAGAGTAATCGCTTGAACCCAG 2038
|||||
DB 114960 CGAGCACCCTGTAAATCCAGCTACTTTGGAGGCTCAGGACGAGAGTAATCGCTTGAACCCAG 115019
|||||
QY 2039 GAGGTGGAGGTGGCACT 2055
|||||
DB 115020 GAGGTGGAGGTGGCACT 115036
|||||
RESULT 56
AC079944 126295 bp DNA linear PRI 21 JUN-2002
LOCUS Homo sapiens 3 BAC RP11-435F17 (Roswell Park Cancer Institute Human
DEFINITION BAC Library) complete sequence.
ACCESSION AC079944
VERSION AC079944.17 GI:21535888
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 126295)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-ouman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleaveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinb,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Fallis,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hum,J.,

```

Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovac, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lounsegh, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenko, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Zaylor, S.L., Weinstein, G. and Gibbs, R.

TITLE JOURNAL

REFERENCE
2 (bases 1 to 126295)

Worley, K.C.

TITLE JOURNAL

Submitted (20-SEP-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 126295)

Worley, K.C.

REFERENCE AUTHORS TITLE JOURNAL

Submitted (20-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 126295)

Worley, K.C.

REFERENCE AUTHORS TITLE JOURNAL

Submitted (21-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jun 21, 2002 this sequence version replaced gi:21465331.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annot.action.html>.

FEATURES

source	Location/Qualifiers
1..126295	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="3"
	/clone="RP11-435F17"
	/complement(7..87)
	/rpt_family="L1ME"
	/complement(1249..1492)
	/rpt_family="MIR"
	2769..2793
	/rpt_family="(T)n"
	/complement(2813..2980)
	/rpt_family="AluSx"
	/complement(3472..3761)
	/rpt_family="AluJo"
	3834..3872
	/rpt_family="AT-rich"
	4088..4194
	/rpt_family="MLT1B"
	4197..4494
	/rpt_family="AluY"
	4496..4634
	/rpt_family="GA-rich"
	4637..4703
	/rpt_family="MLT1B"
	4707..4988
	/rpt_family="AluSx"
	5110..5302
	/rpt_family="MLT1B"
	/complement(5818..6121)
	/rpt_family="AluSg"
	6658..6768
	/rpt_family="MIR"
	7071..7101
	/rpt_family="AT-rich"
	/complement(9057..9110)
	/rpt_family="MIR"
	/complement(9111..10226)
	/rpt_family="Tigger3b"
	10227..10524
	/rpt_family="AluSg"
	/complement(10525..10588)
	/rpt_family="Tigger3b"
	/complement(10589..10749)
	/rpt_family="MIR"
	10793..11077
	/rpt_family="L1MC4"
	11090..11398
	/rpt_family="AluSg"
	11587..11616
	/rpt_family="(GAAA)n"
	12951..13268
	/rpt_family="AluSc"
	13282..13577
	/rpt_family="AluSx"
	13569..13669
	/standard_name="176498"
	/complement(13689..13785)
	/rpt_family="MIR"
	/complement(14091..14393)
	/rpt_family="L1MB7"
	/complement(14650..14704)
	/rpt_family="L2"
	15241..15428
	/rpt_family="MIR"

```

repeat_region 15498..15786
/rpt_family="AluSc"
repeat_region 15787..15813
/rpt_family="TAA)n"
repeat_region complement(15848..16146)
/rpt_family="AluSx"
repeat_region 16268..16323
/rpt_family="AT-rich"
repeat_region complement(16639..16778)
/rpt_family="L2"
repeat_region 17560..17580
/rpt_family="AT-rich"
repeat_region complement(19788..20005)
/rpt_family="L2"
repeat_region complement(20100..20384)
/rpt_family="L2"
repeat_region 21645..21936
/rpt_family="AluSx"
repeat_region 22446..22536
/rpt_family="MER94"
repeat_region complement(22567..22692)
/rpt_family="L2"
repeat_region complement(23033..23341)
/rpt_family="MER58B"
repeat_region 23728..23752

Query Match 2.7% Score 77; DB 9; Length 126295;
Best Local Similarity 100.0%; Pred. No. 1e-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGGCACCTGTAATCCAGCTACTGGAGGCTGAGCAGGAGAGTAATCGCTTGAACCCAG 2038
|||||
DB 109614 GCAGGCACCTGTAATCCAGCTACTGGAGGCTGAGCAGGAGAGTAATCGCTTGAACCCAG 109673
|||||

QY 2039 GAGGTGGAGGTGCAGT 2055
|||||
DB 109674 GAGGTGGAGGTGCAGT 109690

RESULT 57
AC078928
LOCUS AC078928 154968 bp DNA linear HTG 05-SEP-2000
DEFINITION Homo sapiens chromosome 12q clone RP11-22G4, WORKING DRAFT
SEQUENCE, 10 unordered pieces.
AC078928
VERSION AC078928.6 GI:9966147
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,D., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Ducan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Furum-Tansey,J., Frantz,P., Ganesha,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hognes,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichter,A.O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu.L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugang,R.,
Tabor,T., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.
Direct Submission
Unpublished

REFERENCE
AUTHORS Worley,K.C.
TITLE Submitted (11-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 4, 2000 this sequence version replaced gi:9929481.
COMMENT
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCAS
Center clone name: RP11-22G4
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-Primer Bodipy: 80% of reads
Chemistry: Dye-terminator Big Dye: 20% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 150687 bases at least Q40
Consensus quality: 154338 bases at least Q30
Consensus quality: 156644 bases at least Q20
Estimated insert size: 152379; sum-of-ctnigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 5x in Q20 bases; sum-of-ctnigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 10 ctnigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the ctnigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 1 51941: contig of 51941 bp in length
* 51942 52041: gap of unknown length
* 52042 85871: contig of 33830 bp in length
* 85872 85971: gap of unknown length
* 85972 99094: contig of 13123 bp in length
* 99095 99194: gap of unknown length
* 99195 109291: contig of 10097 bp in length
* 109292 109391: gap of unknown length
* 109392 118498: contig of 9107 bp in length
* 118499 118598: gap of unknown length
* 118599 127575: contig of 8977 bp in length
* 127576 127675: gap of unknown length
* 127676 135234: contig of 7559 bp in length
* 135235 135334: gap of unknown length
* 135335 142541: contig of 7207 bp in length
* 142542 142641: gap of unknown length
* 142642 150582: contig of 7941 bp in length
* 150583 150682: gap of unknown length
* 150683 154968: contig of 4286 bp in length.
FEATURES
Location/Qualifiers
1..154968
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12q"
/clone="RP11-22G4"
BASE COUNT 48695 a 29406 c 27663 g 48296 t 908 others
ORIGIN
Query Match 2.7% Score 77; DB 2; Length 154968;
Best Local Similarity 100.0%; Pred. No. 1e-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1979 GCAGGCACCTGTAATCCAGCTACTGGAGGCTGAGCAGGAGAGTAATCGCTTGAACCCAG 2038
|||||
DB 126077 GCAGGCACCTGTAATCCAGCTACTGGAGGCTGAGCAGGAGAGTAATCGCTTGAACCCAG 126136
|||||

```



```
Query Match      2.7%; Score 77; DB 2; Length 181183;
Best Local Similarity 100.0%; Pred. No. 1e-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGGCACCTGTAATCCAGCTACTTGGGAGGCTGAGCAGAGAAATCGCTTGAACCCAG 2038
      |
Db 141532 GCAGGCACCTGTAATCCAGCTACTTGGGAGGCTGAGCAGAGAAATCGCTTGAACCCAG 141473
      |

QY 2039 GAGGTGGAGGTTGCAGT 2055
      |
Db 141472 GAGGTGGAGGTTGCAGT 141456
      |

RESULT 61
AC022281
LOCUS      186431 bp DNA linear HTG 24-JAN-2002
DEFINITION Homo sapiens chromosome 10 clone RP11-10E13, WORKING DRAFT
SEQUENCE, 9 unordered pieces.
ACCESSION AC022281
VERSION   1
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN;
          HTGS_CANCELLED.
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     1 (bases 1 to 186431)
JOURNAL   Genome Therapeutics Corporation Sequencing Center: Human Genome
REFERENCE Smith,D.R.
AUTHORS   Genome Therapeutics Corporation Sequencing Center: Human Genome
TITLE     Sequence Data
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 186431)
AUTHORS   Smith,D.R.
TITLE     Direct Submission
JOURNAL   Submitted (28-JAN-2000) Genome Therapeutics Corporation, 100 Beaver
          Street, Waltham, MA 02453, USA
COMMENT   On Mar 21, 2001 this sequence version replaced gi:9929643.
          -----
          Center: Genome Therapeutics Corporation
          Center code: GTC
          Web site: http://www.genomecorp.com/
          Contact: gtc-seqcenter@genomecorp.com
          -----
          Project Information
          Center project name: hgl143
          -----
          Summary Statistics
          Sequencing vector: N/A
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Assembly program: Phrap; version 990315
          Consensus quality: 395389 bases at least Q40
          Consensus quality: 405067 bases at least Q30
          Consensus quality: 410525 bases at least Q20
          Insert size: 185631; sum-of-contigs
          Quality coverage: 6.4x in Q20 bases; sum-of-contigs
          -----
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 9 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          * 124725: contig of 124725 bp in length
          * 124726 124825: gap of unknown length
          * 124826 125924: contig of 1099 bp in length
          * 125925 126024: gap of unknown length
          * 126025 128493: contig of 2469 bp in length
          * 128494 128593: gap of unknown length
          * 128594 156515: contig of 27922 bp in length
          * 156516 156616: gap of unknown length
          * 156616 170011: contig of 13396 bp in length
          * 170012 170111: gap of unknown length
```

```
* 170112 176430: contig of 6319 bp in length
* 176431 176530: gap of unknown length
* 176531 179019: contig of 2489 bp in length
* 179020 179119: gap of unknown length
* 179120 181965: contig of 2846 bp in length
* 181966 182065: gap of unknown length
* 182066 186431: contig of 4366 bp in length.
FEATURES             Location/Qualifiers
     source            1..186431
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="10"
                        /clone="RP11-10E13"
                        /clone_lib="RPC1-11"
                        1..124725
     misc_feature       /note="assembly_name:Contig243
                        clone_end:SP6"
                        124826..125924
                        /note="assembly_name:Contig204"
                        126025..128493
                        /note="assembly_name:Contig233"
                        128594..156515
                        /note="assembly_name:Contig240"
                        156616..170011
                        /note="assembly_name:Contig239"
                        170112..176430
                        /note="assembly_name:Contig235"
                        176531..179019
                        /note="assembly_name:Contig231"
                        179120..181965
                        /note="assembly_name:Contig232"
                        182066..186431
                        /note="assembly_name:Contig236
                        clone_end:T7"
BASE COUNT 48827 a 41798 c 41896 g 53062 t      848 others
ORIGIN
Query Match      2.7%; Score 77; DB 2; Length 186431;
Best Local Similarity 100.0%; Pred. No. 1e-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1979 GCAGGCACCTGTAATCCAGCTACTTGGGAGGCTGAGCAGAGAAATCGCTTGAACCCAG 2038
      |
Db 132433 GCAGGCACCTGTAATCCAGCTACTTGGGAGGCTGAGCAGAGAAATCGCTTGAACCCAG 132492
      |

QY 2039 GAGGTGGAGGTTGCAGT 2055
      |
Db 132493 GAGGTGGAGGTTGCAGT 132509
      |

RESULT 62
AC007345
LOCUS      188863 bp DNA linear PRI 04-NOV-2000
DEFINITION Homo sapiens chromosome 16 clone RP11-466N18, complete sequence.
ACCESSION AC007345
VERSION   AC007345.5 GI:11094677
KEYWORDS  HTG.
SOURCE    Homo sapiens.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 188863)
AUTHORS   DOE Joint Genome Institute.
TITLE     Sequencing of Human Chromosome 16
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 188863)
AUTHORS   Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
          Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
          Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
          Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
          McMurry,K., Han,C. and Deaven,L.
TITLE     Direct Submission
JOURNAL   Submitted (21-APR-1999) Center for Human Genome Studies, DOE Joint
```

Genome Institute, Los Alamos National Laboratory, MS M888, Los Alamos, NM 87545, USA
3 (bases 1 to 188863)
DOE Joint Genome Institute.
Direct Submission
Submitted (04-NOV-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Nov 4, 2000 this sequence version replaced gi:9795553.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

FEATURES
Source
1..188863
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-466N18"
BASE COUNT 57057 a 36866 c 36556 g 58384 t
ORIGIN

Query Match 2.7% Score 77; DB 9; Length 188863;
Best Local Similarity 100.0%; Pred. No. le-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1979 GCAGGCACCTGTAAATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 2038
|||||
DB 185044 GCAGGCACCTGTAAATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 185103
|||||
QY 2039 GAGGTGGAGGTTGCAGT 2055
|||||
DB 185104 GAGGTGGAGGTTGCAGT 185120
|||||

RESULT 63
AP005264/c
LOCUS
DEFINITION Homo sapiens genomic DNA, chromosome 18 clone:RP11-64C12, complete sequence.
ACCESSION AP005264
VERSION AP005264.3 GI:22202843
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-64C12.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
REFERENCE
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (24-MAY-2002) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@psc.riken.go.jp, URL:<http://hgp.gsc.riken.go.jp/>, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Aug 9, 2002 this sequence version replaced gi:21624263.
FEATURES
Source
1..192539
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18p"
/chromosome="18"
/clone="RP11-64C12"
BASE COUNT 51833 a 45020 c 43248 g 52438 t
ORIGIN

Query Match 2.7% Score 77; DB 9; Length 192539;

Best Local Similarity 100.0%; Pred. No. le-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1979 GCAGGCACCTGTAAATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 2038
|||||
DB 95946 GCAGGCACCTGTAAATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 95887
|||||
QY 2039 GAGGTGGAGGTTGCAGT 2055
|||||
DB 95886 GAGGTGGAGGTTGCAGT 95870
|||||

RESULT 64
CNS01DRV
LOCUS
DEFINITION Human chromosome 14 DNA sequence BAC R-1017G21 of library RPC1-11 from chromosome 14 of Homo sapiens (human), complete sequence.
ACCESSION AL118558
VERSION AL118558.6 GI:15485141
KEYWORDS HTG: HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 194871)
Heilig,R., Petit,J.L., Vico,V., Davila,C., Robert,C., Wincker,P., Hottel,P., Cattolico,L., Barbe,V., Pelletier,E., Arligonave,F., Levy,M., Eckenberg,R., Bruls,T., Deberardinis,V., Croud,C., Gypay,G., Saurin,W. and Weissbach,J.
Sequencing of the human chromosome 14
Unpublished
2 (bases 1 to 194871)
Genoscope.
Direct Submission
Submitted (04-SEP-2001) Genoscope - Centre National de Sequenage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
On Sep 6, 2001 this sequence version replaced gi:14571656.
----- Genome Center
Center: Genoscope / Centre National de Sequenage
Center code: GS
Web site: <http://www.genoscope.cns.fr/>
Contact: Seqref@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : C-2017C7
Downstream BAC (overlapping the SP6 end) : R-114H15 (AC AL14224)
----- Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 9.00x in 920 bases; sum-of-contigs

Overall quality chart :
Range : bases
0 :
1 - 9 :
10 - 19 : 1
20 - 29 : 36
30 - 39 : 176
40 - 49 : 2561
50 - 59 : 6688
60 - 69 : 12185
70 - 79 : 30197
80 - 89 : 70225
90 - 99 : 72822

Percentage of bases with a quality value >= 40 : 99 %
FEATURES
Location/Qualifiers
1..194871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-1017G21"


```
STS      /clone_lib="RPC1-11"
53212.   .53394
/note="matching EMBL:D26445
RHdb: RH25507
RHdb: RH25280
dbSTS: STS3660
Identified using the e-PCR software (G. Schuler)"
53216.   .53395
/note="matching EMBL:R42676
RHdb: RH8632
dbSTS: STS13771
Identified using the e-PCR software (G. Schuler)"
53386.   .53543
/note="matching EMBL:G19911
RHdb: RH10711
dbSTS: STS21713
Identified using the e-PCR software (G. Schuler)"
76990.   .77117
/note="matching EMBL:AA433914
RHdb: RH92864
dbSTS: STS65867
Identified using the e-PCR software (G. Schuler)"
92484.   .92604
/note="matching EMBL:H97017
RHdb: RH92831
dbSTS: STS65634
Identified using the e-PCR software (G. Schuler)"
98049.   .98194
/note="matching EMBL:N46343
RHdb: RH45764
dbSTS: STS36832
Identified using the e-PCR software (G. Schuler)"
161688.  .161850
/note="matching EMBL:T86293
RHdb: RH17951
dbSTS: STS13765
Identified using the e-PCR software (G. Schuler)"
178303.  .178552
/note="matching EMBL:T03672
RHdb: RH53557
dbSTS: STS42866
Identified using the e-PCR software (G. Schuler)"
178326.  .178431
/note="matching EMBL:G19930
RHdb: RH10728
dbSTS: STS1440
Identified using the e-PCR software (G. Schuler)"
178392.  .178577
/note="matching EMBL:T15640
RHdb: RH1404
dbSTS: STS25978
Identified using the e-PCR software (G. Schuler)"
178505.  .178605
/note="matching EMBL:D20215
RHdb: RH44085
RHdb: RH1617
RHdb: RH146
dbSTS: STS22994
Identified using the e-PCR software (G. Schuler)"
BASE COUNT 49542 a 46484 c 46814 g 52031 t
ORIGIN

Query Match      2.7%  Score 77;  DB 9;  Length 194871;
Best Local Similarity 100.0%;  Pred. No. 1e-28;
Matches 77;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY  1979  GCAGGCACCTGTATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 2038
      |||||||
DB  50187  GCAGGCACCTGTATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAG 50246
      |||||||
QY  2039  GAGGTGGAGGTTGCAGT 2055
      |||||||
DB  50247  GAGGTGGAGGTTGCAGT 50263
      |||||||

RESULT 65
AF027390/c
LOCUS      Homo sapiens 225432 bp  DNA  linear  PRI 02-APR-1998
DEFINITION Homo sapiens 7q telomere, complete sequence.
ACCESSION  AF027390
VERSION     AF027390.1 GI:3004858
KEYWORDS   HTG.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 225432)
AUTHORS   Chi, H.-C., Saunders, E.H., Buckingham, J.M., Ricke, D.O., Munk, C.C.,
Lobb, R.R., Ueng, S.Y.-J., Mundt, M.O., White, P.S., Tatum, O.L.,
Rietman, H.C. and Moyzis, R.K.
TITLE     DNA Sequence Analysis of the Terminal 226 kb of Human Chromosome 7q
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 225432)
AUTHORS   Lobb, R.R., Ueng, S.Y.-J., Mundt, M.O., Tatum, O.L., Ricke, D.O., Munk, C.C.,
Moyzis, R.K.
TITLE     Direct Submission
JOURNAL   Submitted (30-SEP-1997) Center for Human Genome Studies, Los Alamos
          National Laboratory, MS M888, New Mexico, NM 87545, USA
FEATURES   Source
            1..225432
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="7"
             /map="7q telomere"
             901..1496
              /note="HSAL000641"
              /rpt_family="Alu"
              complement(2341..2376)
              /note="GRAIL 2 good exon ; ORF; frame 0"
              2513..2800
               /note="M32220"
               /rpt_family="retrovirus"
               2659..2997
                /note="LTR2"
                4101..4283
                 /note="HSAL006653"
                 /rpt_family="Alu"
                 complement(4512..4500)
                 /note="MER3"
                 /rpt_family="MER3"
                 8171..8421
                  /note="GRAIL 2 good exon ; ORF; frame 0"
                  8929..9106
                   /note="GRAIL 2 marginal exon ; ORF; frame 1"
                   12304..12466
                    /note="Z69721"
                    /rpt_family="L1"
                    complement(12460..12687)
                    /note="Z69364"
                    /rpt_family="Alu"
                    complement(12621..12766)
                    /note="97% identity -- U66083"
                    /rpt_family="L1"
                    complement(12792..14659)
                    /note="GRAIL 2 excellent exon ; ORF; frame 2"
                    19550..19636
                     /note="80% identity"
                     /db_xref="dbEST:R38951"
                     19655..19824
                      /note="80% identity -- Z77249"
                      19840..19998
                       /note="83% identity"
                       /db_xref="dbEST:W79023"
                       20128..20260
                        /misc_feature
```

```
/note="81% identity -- 277249"
21027..21097
/note="GRAIL 2 excellent exon ; ORF; frame 2"
complement(24077..24350)
/repeat_region
/rpt_family="Alu"
complement(24301..24492)
/note="HSAL000410"
/rpt_family="Alu"
24633..24901
/rpt_family="Alu"
complement(24992..25278)
/note="273359"
/rpt_family="Alu"
complement(25515..25599)
/note="MALR003165"
/rpt_family="THE1"
complement(25777..25901)
/note="MALR00322"
/rpt_family="MSTC"
complement(26808..27050)
/note="X66298"
/rpt_family="Alu"
complement(27009..27125)
/note="HSAL004729"
/rpt_family="Alu"
complement(27361..27462)
/note="GRAIL 2 excellent exon ; ORF; frame 2"
complement(27482..27721)
/note="GRAIL 2 good exon ; ORF; frame 2"
30922..31155
/rpt_family="MER33"
31815..31991
/note="GRAIL 2 good exon ; ORF; frame 1"
complement(31907..34394)
/gene="VIP2R"
complement(join(<31907..32006,34182..34394))
/product="vasoactive intestinal polypeptide receptor 2
precursor"
complement(31907..32006)
/gene="VIP2R"
/note="99% identity -- L40764"
complement(join(<31908..32006,34182..34234))
/gene="VIP2R"
/codon_start=3
/product="vasoactive intestinal polypeptide receptor 2
precursor"
/db_xref="GI:3004859"
/translating="MRTLLPALLTCWLLAPVNSIHPECRFHLEIOEETKCAELLRS
QTEKHK"
complement(33411..33597)
/gene="VIP2R"
/note="100% identity -- 259087"
complement(34182..34232)
/gene="VIP2R"
/note="100% identity -- L40764"
complement(34200..34394)
/gene="VIP2R"
/note="92% identity -- L36566"
34432..34694
/rpt_family="TAR1"
35000..35058
/note="GRAIL 2 excellent exon ; ORF; frame 2"
complement(35325..35445)
/rpt_family="MER42"
complement(36238..36296)
/note="HSAL004342"
/rpt_family="Alu"
complement(37248..37455)
/note="MER4_00001"
/rpt_family="MER4"
```

```
complement(37454..37580)
/note="84% identity"
/db_xref="dbEST:T62489"
complement(37828..38086)
/note="270040"
/rpt_family="MER4"
complement(41427..41642)
/rpt_family="MLT2B2"
complement(42308..42622)
/rpt_family="MLT2B2"
43022..43522
/rpt_family="Tigger2"
43881..44000
/note="GRAIL 2 good exon ; ORF; frame 2"
44984..45116
/note="GRAIL 2 good exon ; ORF; frame 1"
complement(46422..46478)
/note="GRAIL 2 good exon ; ORF; frame 0"
complement(46801..47121)
/note="HSAL005507"
/rpt_family="Alu"
complement(47511..47779)
/note="GRAIL 2 good exon ; ORF; frame 1"
complement(48123..48196)
/note="MALR00186"
/rpt_family="MLT2"
complement(49320..49497)
/note="GRAIL 2 excellent exon ; ORF; frame 2"
complement(49934..50215)
/note="GRAIL 2 good exon ; ORF; frame 2"
complement(51833..51901)
/note="GRAIL 2 good exon ; ORF; frame 2"
complement(52214..52500)
/note="268162"
/rpt_family="Alu"
54960..55032
/rpt_family="LIMC2"
55462..55537
/rpt_family="LIME2"
57335..58234
/rpt_family="LJ"
complement(58350..58466)
/rpt_family="MER44"
58503..58646
/rpt_family="LIMA9"
complement(58757..59020)
/note="GRAIL 2 good exon ; ORF; frame 0"
62145..62345
/note="GRAIL 2 good exon ; ORF; frame 2"
complement(62518..64744)
/rpt_family="Tigger1"
complement(66883..67070)
/note="GRAIL 2 excellent exon ; ORF; frame 0"
complement(67293..67336)
/note="GRAIL 2 good exon ; ORF; frame 1"
67421..67746
/rpt_family="Alu"
complement(68255..68320)

Query Match 2.7% Score 77; DB 9; Length 225412;
Best Local Similarity 100.0%; Pred. No. le-28;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1979 GCAGGCACCTGTATCCAGCTACTTGGAGGCTGAGGCAGGAGATCGTTGAACCCAG 2038
|||||
Db 124031 GCAGGCACCTGTATCCAGCTACTTGGAGGCTGAGGCAGGAGATCGTTGAACCCAG 2038
|||||

Qy 2039 GAGGTGAGGTTGCAGT 2055
|||||
Db 123971 GAGGTGAGGTTGCAGT 123955

RESULT 66
```

Db	27812	CAGGCACCTGTATATCCAGCTACTTGGAGCGCTGAGCAGGAGAAATCGCTTGAACCCAGG	27871
QY	2040	AGTGGAGGTTGCAGT 2055	
Db	27872	AGTGGAGGTTGCAGT 27887	
RESULT 67			
AC100787/c			
LOCUS	AC100787		
DEFINITION	Homo sapiens chromosome 17 clone CTD-2589J4 map 17, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.		
ACCESSION	AC100787		
VERSION	AC100787.2 GI:22325312		
KEYWORDS	HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 199725)		
AUTHORS	Birren,B., Nusbaum,C. and Lander,E.		
TITLE	Homo sapiens chromosome 17, clone CTD-2589J4		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 199725)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viell,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
REFERENCE	3 (bases 1 to 199725)		
AUTHORS	Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gird,S., Gird,S., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viell,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
COMMENT	On Aug 20, 2002 this sequence version replaced gi:17048157. All repeats were identified using RepeatMasker:		

AC011933	153300 bp	DNA	linear	HTG 31-DEC-2001				
LOCUS	Homo sapiens clone RP11-16C1, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.							
AC011933	AC011933.13 GI:18001668							
VERSION	HTG: HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.							
KEYWORDS	Homo sapiens.							
SOURCE	Homo sapiens							
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
REFERENCE	1 (bases 1 to 153300)							
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.							
TITLE	Homo sapiens, clone RP11-16C1							
JOURNAL	Unpublished							
REFERENCE	2 (bases 1 to 153300)							
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donellan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Raley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.							
TITLE	Direct Submission							
JOURNAL	Submitted (16-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA							
COMMENT	On Dec 31, 2001 this sequence version replaced gi:15451715. All repeats were identified using RepeatMasker: http://ftp.genome.washington.edu/RM/RepeatMasker.html Smit, A.F.A. & Green, P. (1996-1997)							
----- Genome Center -----								
Center: Whitehead Institute/ MIT Center for Genome Research								
Center code: WIBR								
Web site: http://www-seq.wi.mit.edu								
Contact: sequence_submissions@genome.wi.mit.edu								
----- Project Information -----								
Center project name: L3514								
Center clone name: 16_C_1								

* NOTE: This is a 'working draft' sequence. It currently								
* consists of 2 contigs. Gaps between the contigs								
* are represented as runs of N. The order of the pieces								
* is believed to be correct as given, however the sizes								
* of the gaps between them are based on estimates that have								
* provided by the submittor.								
* This sequence will be replaced								
* by the finished sequence as soon as it is available and								
* the accession number will be preserved.								
* 1 115860: contig of 115860 bp in length								
* 115861 115960: gap of 100 bp								
* 115961 153300: contig of 37340 bp in length.								
* Location/Qualifiers								
1. .153300								
/organism="Homo sapiens"								
/db_xref="taxon:9606"								
/clone="RP11-16C1"								
/clone_lib="RPC1-11 Human Male BAC"								
39568 a 34876 c 36023 g 42733 t 100 others								
BASE COUNT								
ORIGIN								
Query Match 2.7%; Score 76; DB 2; Length 153300;								
Best Local Similarity 100.0%; Pred. No. 3.4e-28;								
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;								
QY	1980 CAGGCACCTGTATATCCAGCTACTTGGAGCGCTGAGCAGGAGAAATCGCTTGAACCCAGG 2039							

```

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20503
Center clone name: 2589_J_4
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 39560: contig of 39560 bp in length
* 39561 39660: gap of 100 bp
* 39661 199725: contig of 160065 bp in length.
*
Location/Qualifiers
1..199725
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="CTD-2589J4"
/clone_lib="CIND2 Human BAC"
BASE COUNT 49522 a 50529 c 51605 g 47969 t 100 others
ORIGIN

Query Match 2.7%; Score 76; DB 2; Length 199725;
Best Local Similarity 100.0%; Pred. No. 3.4e-28;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1980 CAGGCACCTGTATATCCAGCTACTTGGAGGTGAGGAGGAAATCGTTGAACCCAGG 2039
|||||
50275 CAGGCACCTGTATATCCAGCTACTTGGAGGTGAGGAGGAAATCGTTGAACCCAGG 50216

QY 2040 AGTGGAGCTTGCAGT 2055
|||||
50215 AGTGGAGCTTGCAGT 50200

RESULT 68
AC024328
LOCUS AC024328 66452 bp DNA linear HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-30023, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC024328
VERSION AC024328.1 GI:7108083
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 66452)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome, clone RP11-30023
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 66452)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeAreltano,K., Dewar,K.P., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fensteror,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczyk,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,

```

TITLE JOURNAL COMMENT

```

McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Mceldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4912
Center clone name: 30_O_23
-----
* NOTE: This record contains 84 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 658: contig of 658 bp in length
* 659 758: gap of 100 bp
* 759 1471: contig of 713 bp in length
* 1472 1571: gap of 100 bp
* 1572 2278: contig of 707 bp in length
* 2279 2378: gap of 100 bp
* 2379 3055: contig of 677 bp in length
* 3056 3155: gap of 100 bp
* 3156 3829: contig of 674 bp in length
* 3830 3929: gap of 100 bp
* 3930 4610: contig of 681 bp in length
* 4611 4710: gap of 100 bp
* 4711 5398: contig of 688 bp in length
* 5399 5498: gap of 100 bp
* 5499 6192: contig of 694 bp in length
* 6193 6292: gap of 100 bp
* 6293 6989: contig of 697 bp in length
* 6990 7089: gap of 100 bp
* 7090 7789: contig of 700 bp in length
* 7790 7889: gap of 100 bp
* 7890 8595: contig of 706 bp in length
* 8596 8695: gap of 100 bp
* 8696 9383: contig of 688 bp in length
* 9384 9483: gap of 100 bp
* 9484 10189: contig of 706 bp in length
* 10190 10289: gap of 100 bp
* 10290 10972: contig of 683 bp in length
* 10973 11072: gap of 100 bp
* 11073 11744: contig of 672 bp in length
* 11745 11844: gap of 100 bp
* 11845 12507: contig of 663 bp in length
* 12508 12607: gap of 100 bp
* 12608 13302: contig of 695 bp in length
* 13303 13402: gap of 100 bp
* 13403 14089: contig of 687 bp in length
* 14090 14189: gap of 100 bp
* 14190 14894: contig of 705 bp in length

```

* 14895 14994: gap of 100 bp
* 14995 15703: contig of 709 bp in length
* 15704 15803: gap of 100 bp
* 15804 16511: contig of 708 bp in length
* 16512 16611: gap of 100 bp
* 16612 17284: contig of 673 bp in length
* 17285 17384: gap of 100 bp
* 17385 18098: contig of 714 bp in length
* 18099 18198: gap of 100 bp
* 18199 18905: contig of 707 bp in length
* 18906 19005: gap of 100 bp
* 19006 19710: contig of 705 bp in length
* 19711 19810: gap of 100 bp
* 19811 20486: contig of 676 bp in length
* 20487 20586: gap of 100 bp
* 20587 21266: contig of 680 bp in length
* 21267 21366: gap of 100 bp
* 21367 22034: contig of 668 bp in length
* 22035 22134: gap of 100 bp
* 22135 22814: contig of 680 bp in length
* 22815 22914: gap of 100 bp
* 22915 23594: contig of 680 bp in length
* 23595 23694: gap of 100 bp
* 23695 24383: contig of 689 bp in length
* 24384 24483: gap of 100 bp
* 24484 25187: contig of 704 bp in length
* 25188 25287: gap of 100 bp
* 25288 25989: contig of 702 bp in length
* 25990 26089: gap of 100 bp
* 26090 26798: contig of 709 bp in length
* 26799 26998: gap of 100 bp
* 26999 27607: contig of 709 bp in length
* 27608 27707: gap of 100 bp
* 27708 28387: contig of 680 bp in length
* 28388 28487: gap of 100 bp
* 28488 29152: contig of 665 bp in length
* 29153 29252: gap of 100 bp
* 29253 29944: contig of 692 bp in length
* 29945 30044: gap of 100 bp
* 30045 30733: contig of 689 bp in length
* 30734 30833: gap of 100 bp
* 30834 31548: contig of 715 bp in length
* 31549 31648: gap of 100 bp
* 31649 32356: contig of 708 bp in length
* 32357 32456: gap of 100 bp
* 32457 33164: contig of 708 bp in length
* 33165 33264: gap of 100 bp
* 33265 33938: contig of 674 bp in length
* 33939 34038: gap of 100 bp
* 34039 34751: contig of 713 bp in length
* 34752 34851: gap of 100 bp
* 34852 35561: contig of 710 bp in length
* 35562 35661: gap of 100 bp
* 35662 36313: contig of 652 bp in length
* 36314 36413: gap of 100 bp
* 36414 37093: contig of 680 bp in length
* 37094 37193: gap of 100 bp
* 37194 37868: contig of 675 bp in length
* 37869 37968: gap of 100 bp
* 37969 38647: contig of 679 bp in length
* 38648 38747: gap of 100 bp
* 38748 39435: contig of 688 bp in length
* 39436 39535: gap of 100 bp
* 39536 40233: contig of 698 bp in length
* 40234 40333: gap of 100 bp
* 40334 41023: contig of 690 bp in length
* 41024 41123: gap of 100 bp
* 41124 41824: contig of 701 bp in length
* 41825 41924: gap of 100 bp
* 41925 42607: contig of 683 bp in length
* 42608 42707: gap of 100 bp
* 42708 43411: contig of 704 bp in length
* 43412 43511: gap of 100 bp

* 43512 44217: contig of 706 bp in length
* 44218 44317: gap of 100 bp
* 44318 45001: contig of 684 bp in length
* 45002 45101: gap of 100 bp
* 45102 45783: contig of 682 bp in length
* 45784 45883: gap of 100 bp
* 45884 46575: contig of 692 bp in length
* 46576 46675: gap of 100 bp
* 46676 47368: contig of 693 bp in length
* 47369 47468: gap of 100 bp
* 47469 48149: contig of 681 bp in length
* 48150 48249: gap of 100 bp
* 48250 48949: contig of 700 bp in length
* 48950 49049: gap of 100 bp
* 49050 49756: contig of 707 bp in length
* 49757 49856: gap of 100 bp
* 49857 50557: contig of 701 bp in length
* 50558 50657: gap of 100 bp
* 50658 51340: contig of 683 bp in length
* 51341 51440: gap of 100 bp
* 51441 52140: contig of 700 bp in length
* 52141 52240: gap of 100 bp
* 52241 52932: contig of 692 bp in length
* 52933 53032: gap of 100 bp
* 53033 53719: contig of 687 bp in length
* 53720 53819: gap of 100 bp
* 53820 54503: contig of 684 bp in length
* 54504 54603: gap of 100 bp
* 54604 55294: contig of 691 bp in length

Query Match 2.6%; Score 74; DB 2: Length 66452;
Best Local Similarity 100.0%; Pred. No. 3.8e-27;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTGTATCCAGCTACTTGGAGGCTGAGCAGAGAAATCGCTTGAACCCAGGAG 2041
|||||
Db 9693 GGCACCTGTATCCAGCTACTTGGAGGCTGAGCAGAGAAATCGCTTGAACCCAGGAG 9752
|||||
QY 2042 GTGGAGGTGCAGT 2055
|||||
Db 9753 GTGGAGGTGCAGT 9766

RESULT 69
AC008754/c
LOCUS AC008754 66792 bp DNA linear PRI 28-FEB-2001
DEFINITION Homo sapiens chromosome 19 clone CTD-3023J11, complete sequence.
ACCESSION AC008754
VERSION AC008754.8 GI:13162498
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 66792)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 66792)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 66792)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Feb 28, 2001 this sequence version replaced gi:9256040.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www-shgc.stanford.edu

Quality: Phrap Quality >=40 99.6% of Sequence;
Estimated Total Number of Errors is 0.2.

```

FEATURES
Source
1..66792
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTD-3023J11"
BASE COUNT      16364 a 17164 c 17596 g 15668 t
ORIGIN
Query Watch          2.6%; Score 74; DB 9; Length 66792;
Best Local Similarity 100.0%; Pred. No. 3.8e-27;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1982 GGCACCTGTAATCCAGCTACTTGGAGGCTGAGCAGGAGTAATCGCTTGAACCCAGGAG 2041
|||||
DB 15755 GGCACCTGTAATCCAGCTACTTGGAGGCTGAGCAGGAGTAATCGCTTGAACCCAGGAG 15696
QY 2042 GTGGAGGTTGCAGT 2055
|||||
DB 15695 GTGGAGGTTGCAGT 15682

```

```

RESULT 70
AC092579
LOCUS      118958 bp      DNA      linear      PRI 21-FEB-2002
DEFINITION Homo sapiens BAC clone RP11-15K19 from 7, complete sequence.
ACCESSION  AC092579
VERSION    AC092579.3 GI:18653765
KEYWORDS  HTG.
SOURCE     Homo sapiens.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 118958)
Sulston,J.B. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
2 (bases 1 to 118958)
Cordum,H., Elliott,G., Spalding,L. and Phillips,A.
The sequence of Homo sapiens BAC clone RP11-15K19
Unpublished (2001)
3 (bases 1 to 118958)
Waterston,R.H.
Direct Submission
Submitted (19-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 118958)
Waterston,R.H.
Direct Submission
Submitted (13-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 118958)
Waterston,R.
Direct Submission
Submitted (21-FEB-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 13, 2002 this sequence version replaced gi:18425307.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0015K19
Drafting Center: WIBR
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, H., Frenken, E., Taten, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-419M24, 2000 bp overlap; the clone sequenced to the right is RP11-745J15. Actual start of this clone is at base position 160861 of RP11-419M24; actual end is at base position 118958 of RP11-15K19.

Data from AC092101 was used to finish AC092579.

The sequence of AC011922 has been incorporated into AC092579.

FEATURES	Location/Qualifiers
source	1..118958 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /map="7"
repeat_region	/clone="RP11-15K19" 1..139 /rpt_family="Alu"
repeat_region	193..497 /rpt_family="Alu"
misc_feature	228..245 /note="match to EST W02526 (NID:g1274505) z148e05.r1"
repeat_region	553..864 /rpt_family="Alu"
misc_feature	821..1090 /note="match to EST BF980347 (NID:g12347562)"
repeat_region	918..993 /rpt_family="MER1_type?"
misc_feature	1053..1499 /note="match to EST BF925611 (NID:g12322107)"
repeat_region	1646..1957 /rpt_family="Alu"
repeat_region	2002..2270 /rpt_family="MaLR"
repeat_region	2271..3282 /rpt_family="ERVK"
misc_feature	2920..2928 /note="match to EST AA922972 (NID:g3070281) ok77d11.s1"
repeat_region	3283..3428

```
/rpt_family="MaLR"
3503..3812
/rpt_family="Alu"
4056
/Note="match to EST AW903301 (NID:g8067506)"
4086..4447
/Note="match to EST AW903301 (NID:g8067506)"
4417..4720
/rpt_family="Alu"
4574..4575
/Note="match to EST AW903301 (NID:g8067506)"
4790..5093
/rpt_family="Alu"
5248..5250
/Note="match to EST N68717 (NID:g1224878) za20a06.sl"
5376..5830
/Note="match to EST AI492948 (NID:g4393951) qz47a04.xl"
5933..6286
/rpt_family="MaLR"
6328..6602
/rpt_family="Alu"
6340..6369
/Note="match to EST N68717 (NID:g1224878) za20a06.sl"
6988..7281
/rpt_family="Alu"
7282..7543
/rpt_family="Alu"
7443..7469
/Note="match to EST BF668855 (NID:g11942750)"
8243..8540
/rpt_family="Alu"
8541..8900
/rpt_family="LI"
8898..8915
/Note="similar to Homo sapiens EST B1856928
(NID:g15997675)"
8901..9212
/rpt_family="Alu"
9213..9329
/rpt_family="LI"
9330..9601
/rpt_family="Alu"
9602..9904
/rpt_family="LI"
9905..10191
/rpt_family="Alu"
10192..10412
/rpt_family="LI"
10413..10718
/rpt_family="Alu"
10727..11036
/rpt_family="Alu"
11058..12390
/Note="CpG island (GC=60.2, o/e=0.80, #CpGs=97)"
11260..11309
/Note="match to EST B1458323 (NID:g15248979)"
11280..11309
/Note="match to EST BE827799 (NID:g10260177)"
11281..11309
/Note="match to EST BE439838 (NID:g9439449)"
11297..11309
/Note="match to EST AA358556 (NID:g2010873)"
11341..11655
/Note="match to EST BF237438 (NID:g11151356)"
11678..11750
/rpt_family="MTR"
11830..12121
/rpt_family="Alu"
12044..12049
/Note="match to EST N55773 (NID:g1198621)"
12129..12428
/rpt_family="Alu"
12550..12697

misc_feature
/Note="match to EST AL539603 (NID:g12868971)"
12558..12697
/Note="match to EST AA358556 (NID:g2010873)"
12558..12697
/Note="match to EST BE439838 (NID:g9439449)"
12558..12697
/Note="match to EST BE827799 (NID:g10260177)"

Query Match
Best Local Similarity 2.6%; Score 74; DB 9; Length 118958;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTGTAATCCAGCTACTTGGGAGGCTGAGCAGAGAGTAATCGCTTGACCCAGGAG 2041
Db 60317 GGCACCTGTAATCCAGCTACTTGGGAGGCTGAGCAGAGAGTAATCGCTTGACCCAGGAG 60376
QY 2042 GTGGAGGTTGCAGT 2055
Db 60377 GTGGAGGTTGCAGT 60390

RESULT 71
AF214635 158213 bp DNA linear HTG 26-JUL-2002
LOCUS Homo sapiens chromosome 8 clone XX-CTB788C1, WORKING DRAFT
DEFINITION SEQUENCE, 9 unordered pieces.
ACCESSION AF214635
VERSION AF214635.3 GI:14280180
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
AUTHORS Reichwald,K., Baumgart,C., Blechschmidt,K., Dette,M., Jahn,N.,
Lehmann,R., Menzel,U., Polley,A., Schilhabel,M.B., Schudy,A.,
Siddiqui,R., Taudien,S., Wen,G., Korenberg,J.R., Rosenthal,A. and
Platzter,M.
1 (bases 1 to 158213)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
Chromosome 8 genomic sequence
Unpublished
2 (bases 1 to 158213)
Reichwald,K., Blechschmidt,K., Menzel,U., Baumgart,C., Dette,M.,
Jahn,N., Schilhabel,M., Korenberg,J.R. and Rosenthal,A.
Direct Submission
TITLE
JOURNAL
COMMENT
On Jun 2, 2001 this sequence version replaced gi:8152159.
----- Genome Center
Center: Institute of Molecular Biotechnology
Center code: IMB
Web site: http://genome.imb-jena.de/
Contact: gscj-submit@genome.imb-jena.de
----- Project Information
Center project name: H195
Center clone name: XX-CTB788C1
----- Summary Statistics
Sequencing vector: M13; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 146316 bases at least Q40
Consensus quality: 151479 bases at least Q30
Consensus quality: 154827 bases at least Q20
Quality coverage: 5.51 x in Q20 bases; sum-of-contigs
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
```

```
*****
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3530: contig of 3530 bp in length
* 3531 3330: gap of unknown length
* 3631 5131: contig of 1501 bp in length
* 5132 5231: gap of unknown length
* 5232 13445: contig of 8214 bp in length
* 13446 13545: gap of unknown length
* 13546 31026: contig of 17481 bp in length
* 31027 31126: gap of unknown length
* 31127 51905: contig of 20779 bp in length
* 51906 52005: gap of unknown length
* 52006 81566: contig of 29561 bp in length
* 81567 81666: gap of unknown length
* 81667 111349: contig of 29683 bp in length
* 111350 111449: gap of unknown length
* 111450 154738: contig of 43289 bp in length
* 154739 154838: gap of unknown length
* 154839 158213: contig of 3375 bp in length.
FEATURES             Location/Qualifiers
     source            1..158213
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="XX-CTB788C1"
     misc_feature      1..4
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature      158210..158213
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT  42142 a 33754 c 33773 g 47744 t      800 others
ORIGIN
Query Match      2.6%; Score 74; DB 2: Length 158213;
Best Local Similarity 100.0%; Pred. No. 3.7e-27;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1986 CCGTGAATCCCGAGCTACTGGGAGCGCTGAGCAGGAGGAATCGCTGAACCCAGGAGGTGG 2045
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 137693 CCGTGAATCCCGAGCTACTGGGAGCGCTGAGCAGGAGGAATCGCTGAACCCAGGAGGTGG 137752
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2046 AGGTTGCAGTAAGC 2059
      |||||||||||||||
Db 137753 AGGTTGCAGTAAGC 137766
      |||||||||||||||
RESULT 72
AC105081/c 158605 bp DNA linear PRI 27-APR-2002
LOCUS Homo sapiens chromosome 8, clone CTC-788C1, complete sequence.
DEFINITION AC105081
ACCESSION AC105081
VERSION AC105081.4 GI:20258535
KEYWORDS HTG.
SOURCE
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; vertebrata: Euteleostomi;
          Mammalia; Eutheria; primates; Catarrhini: Hominidae; Homo.
          1 (bases 1 to 158605)
REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
AUTHORS Homo sapiens chromosome 8, clone CTC-788C1
TITLE Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 158605)
          Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
          Anderson,S., Barna,N., Bastien,V., Boquslavsky,L., Boukhgalter,B.,
          Brown,A., Camarata,A., Chang,J., Chazaro,B.,
          Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
          Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
          Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
          Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
          Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
          Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K., Lamazars,R.,
          Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
          MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
          McCarthy,M., McEwan,P., McKernan,K., Meidrim,J., Menius,C.,
          Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
          Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
          Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
          Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
          Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
          Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
          Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
          Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
          Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
          Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
```

```
Anderson,S., Barna,N., Bastien,V., Boquslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazars,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meidrim,J., Menius,C.,
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 158605)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boquslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,A., Campopiano,A., Chang,J.,
Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collamore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazars,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meidrim,J., Menius,C.,
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 158605)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boquslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,A., Campopiano,A., Chang,J.,
Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collamore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazars,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meidrim,J., Menius,C.,
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
```


TITLE Direct Submission
 JOURNAL Submitted (27-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On Apr 21, 2002 this sequence version replaced gi:20128267.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RW/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L21405
 Center clone name: 788_C_1

FEATURES

Source

Location/Qualifiers

1..158605
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="CTC-788C1"
 /clone_lib="CIRC Human BAC"
 complement(339..543)
 /rpt_family="MLT1k"
 545..709
 /rpt_family="GA-rich"
 complement(1479..1582)
 /rpt_family="MIR"
 1637..1664
 /rpt_family="AT-rich"
 complement(4850..5036)
 /rpt_family="MER5A"
 5091..5119
 /rpt_family="(A)n"
 5290..5447
 /rpt_family="LIMA3"
 complement(5645..5710)
 /rpt_family="MER91A"
 complement(5740..5803)
 /rpt_family="MER91A"
 6092..6264
 /rpt_family="MLT1h"
 6360..7020
 /rpt_family="MER67B"
 7101..7221
 /rpt_family="MLT1h1"
 complement(7495..7856)
 /rpt_family="MLT1A1"
 complement(8994..9054)
 /rpt_family="L2"
 complement(9121..9220)
 /rpt_family="MIR"
 complement(9353..9430)
 /rpt_family="L2"
 complement(9454..9527)
 /rpt_family="MIR"
 10413..10430
 /rpt_family="MIR"
 complement(10431..10651)
 /rpt_family="MER58A"
 10652..10879
 /rpt_family="MIR"
 12388..12447
 /rpt_family="L2"
 12805..12827
 /rpt_family="MIR"
 13849..14151
 /rpt_family="MSTD"
 14152..14451
 /rpt_family="AluSx"
 14452..14520

repeat_region /rpt_family="MSTD"
 14546..14601
 /rpt_family="HAL1"
 repeat_region 14603..15150
 /rpt_family="HAL1"
 repeat_region 15148..15313
 /rpt_family="LIMB2"
 repeat_region 15314..15616
 /rpt_family="AluSx"
 repeat_region 15617..15661
 /rpt_family="LIMB2"
 repeat_region 15662..15693
 /rpt_family="(TG)n"
 repeat_region 15694..15729
 /rpt_family="LIMB2"
 repeat_region 15730..15759
 /rpt_family="(GTGTG)n"
 repeat_region 15760..15849
 /rpt_family="LIMB2"
 repeat_region 15850..15974
 /rpt_family="AluY"
 repeat_region 15975..16001
 /rpt_family="(CAAAA)n"
 repeat_region 16002..16179
 /rpt_family="AluY"
 repeat_region 16180..16345
 /rpt_family="LIMB2"
 repeat_region 16346..16750
 /rpt_family="LIPA2"
 repeat_region 16819..16940
 /rpt_family="LIMB2"
 repeat_region 17420..17513
 /rpt_family="LIMB2"
 repeat_region complement(17575..17683)
 /rpt_family="LIMC4"
 repeat_region 17812..17896
 /rpt_family="LIMB2"
 repeat_region 18004..18034
 /rpt_family="AT-rich"

Query Match 2.6%; Score 74; DB 9; Length 158605;

Best Local Similarity 100.0%; Pred.No. 3.7e-27;

Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGTAATCCCGAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGTGG 2045

Db 129206 CCTGTAATCCCGAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAGTGG 129147

QY 2046 AGGTTGCAGTAAGC 2059

Db 129146 AGGTTGCAGTAAGC 129133

RESULT 73

AC068793/c

LOCUS

DEFINITION Homo sapiens chromosome 12 clone RP11-414A12, WORKING DRAFT

SEQUENCE, 8 unordered pieces.

ACCESSION AC068793

VERSION AC068793.19 GI:21431064

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 160371)

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhaý,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratochic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabot,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 160371)
Worley,K.C.

Direct Submission
Submitted (10-MAY-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 160371)
Worley,K.C.

Direct Submission
Submitted (18-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jun 17, 2002 this sequence version replaced gi:20335641.

----- Genome Center of Medicine
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Drafting Center Code: BCM
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HBGG
Center clone name: RP11-414A12
----- Summary Statistics
Sequencing vector: M13;
Chemistry: Dye-primer Bodipy: 19% of reads
Chemistry: Dye-terminator Big Dye: 81% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 153802 bases at least Q40
Consensus quality: 156243 bases at least Q30
Consensus quality: 157405 bases at least Q20
Estimated insert size: 161333; sum-of-ctngs estimation
Quality coverage: 7x in Q20 bases; sum-of-ctngs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).

* NOTE: This is a "working draft" sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1
5965: contig of 5965 bp in length
5966: gap of unknown length
6066: contig of 6296 bp in length
12362: gap of unknown length
12362: gap of unknown length
20343: contig of 7882 bp in length
20344: gap of unknown length
35164: contig of 14720 bp in length
35164: gap of unknown length
35263: contig of 25874 bp in length
61137: gap of unknown length
61138: contig of 25285 bp in length
86522: gap of unknown length
86523: contig of 26271 bp in length
112893: contig of 26271 bp in length
112894: gap of unknown length
112994: contig of 47378 bp in length.

FEATURES

Location/Qualifiers
1..160371
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-414A12"

BASE COUNT 45397 a 33641 c 34490 g 46137 t 706 others
ORIGIN

Query Match 2.6%; Score 74; DB 2: Length 160371;
Best Local Similarity 100.0%; Pred. No. 3.7e-27;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1982 GGCACCTCTAATCCAGCTACTTGGAGGCTGAGCAGGAGAGATCGCTTGAACCCAGGAG 2041
|||||
Db 78319 GGCACCTCTAATCCAGCTACTTGGAGGCTGAGCAGGAGAGATCGCTTGAACCCAGGAG 78260
|||||
Oy 2042 GTGGAGGTTGCAGT 2055
|||||
Db 78259 GTGGAGGTTGCAGT 78246

RESULT 74
AC012156/c
LOCUS
DEFINITION
AC012156 Homo sapiens 162200 bp DNA linear PRI 12-AUG-2000
Institute Human Bac Library (Rosewell Park Cancer complete sequence.)

AC012156
VERSION
KEYWORDS
SOURCE
ORGANISM
HTG.
AC012156.14 GI:7363387
Homo sapiens.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 162200)

REFERENCE
AUTHORS
Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenbiller,K., Bodoty,B., Bouck,J., Bowie,S., Brooks,A., Bulay,C., Bunat,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rasliid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L., Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M., Holloway,C., Hosak,H., Issar,A., Jackson,L.E., Jackson,L., Jia,Y., Jones,M., Kelly,S., Kneitz,S., Kondejewski,N., Kong,Y., Kovar,C., Lau,S., Leal,B., Lee,E., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J., Lucier,R., Marondel,I., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Merscher,S., Miller,A., Montgomery,K.T., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Shim,C., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S., Weinstein,G.,

Weinstock, I.R., Williamson, A., Worley, K., Wren, J., Wrensford, G.,
Xiang, A.M., Yang, R., Yu, W., Zhou, X., Kuchelapatti, R., Nelson, D. and
Gibbs, R.

JOURNAL

REFERENCE

AUTHORS

JOURNAL

2. (bases 1 to 162200)
Worley, K.C.
Direct Submission
Submitted (21-OCT-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

JOURNAL

3. (bases 1 to 162200)
Worley, K.C.
Direct Submission
Submitted (31-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

JOURNAL

4. (bases 1 to 162200)
Worley, K.C.
Direct Submission
Submitted (12-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Mar 31, 2000 this sequence version replaced gi:7211838.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least 2 exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

Summary Statistics

Contig length: 162200
Phrap values in estimate: 156878
Average error rate (BCM-Phrap estimate): 0.000360119
Fraction of Phrap values less than 40 : 0.0298257
Number of consensus changing edits: 53
Number of N's in consensus : 0

Consensus changing edits

Position

Original+Context

Edited+Context

27 atagtactg(t)ttgtgtgca

18532 tatctctatg(t)tttttttt

18533 attctctatg(t)tttttttt

18534 ttctctatg(t)tttttttt

18561 tttttttt(g)gagacagat

18564 tttttttt(g)acagatctt

18568 ttgtgacac(t)agctctgct

18714 agctgattt(t)ttttgatt

18730 gttttttt(g)gagacaggg

18734 ttgtgagga(n)acgctgttc

18754 cactgttta(n)ccagatggt

18781 ctctgacct(n)gtagctgcc

18783 ctgacct(n)gtagctgcc

18795 atccgacct(n)tggtccccc

18797 cgcacct(n)ggccccccaa

22643 tacccttttc(n)acacagcat

22705 cctttctatt(n)tacaaccc

22712 attttacaaa(n)ccgcatgt

22760 tgggtacacc(n)ncagacggg

22761 ggttacacc(n)ccagacggg

34998 aactctggag(n)taagacaaa

35000 ctctggagnt(n)aaagcacaag

35410 gcccttaagt(n)ctgtctgcc

55354 gttttattga(n)aaagcacaag

55718 ctttagatg(n)attttctct

55775 gagaattttc(a)accataaac

55836 tagctgacct(n)attatgctt

55992 cacaagatc(n)ctnncttcat

55995 aaagtacct(n)cttcatatt

55996 aagtacnct(n)cttcatatt

56268 tggtagaacac(n)ngtagtccca

56269 ggtgaacac(n)ngtagtccca

67116 gccagacct(n)taagtaacc

102871 agagattttt(n)ttccacttca

102967 tttttttt(a)aaacagatc

119358 ccaagtatac(n)cttacttgg

131009 tgaataaat(c)caataaaga

133724 tgcacctga(n)tgacctccag

138630 agagagctga(n)gcagcgnnn

138635 gctgagcag(n)gcagcgnnn

138638 gctgagcag(n)gcagcgnnn

138639 angcagcgn(n)nnntcnaga

138640 gcagcgnnn(n)ntcnagacc

138641 gcagcgnnn(n)ntcnagacca

138642 cagcgnnn(n)ntcnagaccag

138645 ncggnnttc(n)agaccagcct

139044 gctctttta(n)aaagaaaga

141300 taaagcaaca(n)taaaagaaa

151715 ctggagtga(n)tggtgngatc

151721 tgaattgtg(n)gactcggct

151751 ctccgctcc(n)aggttcaagc

155920 agtgcggga(n)acagggcgtg

155921 gtcgtgggat(t)acagggcgtg

Distribution of Quality < 40 Bases

# bases	5	10	15	20	25	30	35	40
10001	*	*	*	*	*	*	*	*
9001	*	*	*	*	*	*	*	*
8001	*	*	*	*	*	*	*	*
7001	*	*	*	*	*	*	*	*
6001	*	*	*	*	*	*	*	*
5001	*	*	*	*	*	*	*	*
4001	*	*	*	*	*	*	*	*
3001	*	*	*	*	*	*	*	*
2001	*	*	*	*	*	*	*	*
1001	*	*	*	*	*	*	*	*
01	*	*	*	*	*	*	*	*

Phrap Value Range

```

-----
Version: 1.01 oxfo
Location/Qualifiers
1. .162200
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-433D24"
complement(280..585)
/rpt_family="AluSg"
complement(1291..1341)
/rpt_family="L1MA6"
1342..1641
/rpt_family="AluSx"
complement(1642..1669)
/rpt_family="L1MA6"
1670..1698
/rpt_family="(CA)n"
complement(1699..1882)
/rpt_family="L1MA6"
4674..4845
/rpt_family="MER105"
5070..5848
/rpt_family="L1PA3"

Query Match 2.6% Score 74; DB 9; Length 162200;
Best Local Similarity 100.0%; Pred. No. 3.7e-27;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1982 GGCACCTGTATCCAGCTACTTGGGAGGTGAGGAGGAGAAATCGCTTGAACCCAGGAG 2041
Db 155789 GGCACCTGTATCCAGCTACTTGGGAGGTGAGGAGGAGAAATCGCTTGAACCCAGGAG 155730
Qy 2042 GTGGAGGTTGCAGT 2055
Db 155729 GTGGAGGTTGCAGT 155716

RESULT 75
AC073548 167722 bp DNA linear PRI 31-JUL-2002
LOCUS Homo sapiens chromosome 19 clone RP11-43N16, complete sequence.
DEFINITION Homo sapiens
ACCESSION AC073548
VERSION AC073548.5 GI:22024556
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 167722)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 167722)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 167722)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 31, 2002 this sequence version replaced gi:15022024.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.5% of Sequence;
Estimated Total Number of Errors is 0.7.
NOTE: Forced join 65004.
Location/Qualifiers
-----
Version: 1.01 oxfo
Location/Qualifiers
1. .167722
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-43N16"
65004
/note="NOTE: Forced join 65004"

BASE COUNT 38329 a 46532 c 43717 g 39144 t
ORIGIN

Query Match 2.6% Score 74; DB 9; Length 167722;
Best Local Similarity 100.0%; Pred. No. 3.7e-27;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1982 GGCACCTGTATCCAGCTACTTGGGAGGTGAGGAGGAGAAATCGCTTGAACCCAGGAG 2041
Db 32512 GGCACCTGTATCCAGCTACTTGGGAGGTGAGGAGGAGAAATCGCTTGAACCCAGGAG 32571
Qy 2042 GTGGAGGTTGCAGT 2055
Db 32572 GTGGAGGTTGCAGT 32585

RESULT 76
AC018633 169234 bp DNA linear PRI 21-JAN-2000
LOCUS Homo sapiens clone UWGC:djs1 or RP11-16G1 from 7p14-15, complete
sequence.
DEFINITION Homo sapiens
ACCESSION AC018633
VERSION AC018633.2 GI:6729063
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169234)
AUTHORS Kaul,R.K., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D., Green,P.
and Olson,M.V.
TITLE Large-scale MCD Mapping and Sequencing of Human Chromosome 7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 169234)
AUTHORS Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1999) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE 3 (bases 1 to 169234)
AUTHORS Kaul,R.K. and Richards,B.K.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-2000) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
COMMENT On Jan 21, 2000 this sequence version replaced gi:6579285.
-----
Center: University of Washington Genome Center
Center code: UWGC
Web site: http://genome.washington.edu
Contact: uwgchelp@u.washington.edu
----- Project Information
Center project name: chr-7
Center clone name: djs1 (RP11-16G1)
----- Summary Statistics
Sequencing vector: M13; 100% of reads
Chemistry: Dye-Primer Body; 90% of reads
Chemistry: Dye-terminator Big Dye; 10% of reads
Assembly program: Phrap; version 0.990319
Insert size: 169 234; sum-of-contents
Quality coverage: 11.1X in Q20 bases; sum-of-contents
-----
Overlapping Sequences:
5': UWGC:djs21
3': UWGC:djs29
-----
Sequence Quality Assessment:

```

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Double stranded (DS) coverage: 96.4%
DS or two chemistry coverage: 97.1%
Single stranded regions: 10

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-600 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BglII	FP	Seq	FP	Seq	EcoRI	FP	Seq	FP	Seq	NsiI
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
2055.00	2061.00	8954.00	8687.00	11080.00	11109.00	-----	-----	-----	-----	-----
5487.00	5270.00	7648.00	7660.00	4611.00	4562.00	-----	-----	-----	-----	-----
1913.00	1893.00	7648.00	7556.00	11080.00	11183.00	-----	-----	-----	-----	-----
1508.00	1491.00	838.00	801.00	1377.00	1369.00	-----	-----	-----	-----	-----
12901.00	13322.00	4399.00	4303.00	2193.00	2193.00	-----	-----	-----	-----	-----
1004.00	992.00	4140.00	4094.00	3735.00	3692.00	-----	-----	-----	-----	-----
529.00	503.00	660.00	643.00	614.00	599.00	-----	-----	-----	-----	-----
10358.00	10526.00	3053.00	3100.00	3918.00	3851.00	-----	-----	-----	-----	-----
3805.00	3819.00	5075.00	5106.00	614.00	593.00	-----	-----	-----	-----	-----
1674.00	1645.00	6909.00	6771.00	2570.00	2518.00	-----	-----	-----	-----	-----
2228.00	2170.00	1911.00	1874.00	1997.00	1995.00	-----	-----	-----	-----	-----
3051.00	2974.00	3635.00	3584.00	6199.00	6176.00	-----	-----	-----	-----	-----
14433.00	14779.00	3975.00	3996.00	4913.00	5009.00	-----	-----	-----	-----	-----
1508.00	1475.00	897.00	864.00	2193.00	2184.00	-----	-----	-----	-----	-----
1865.00	1874.00	1410.00	1399.00	12585.00	12586.00	-----	-----	-----	-----	-----
2228.00	2269.00	5716.00	5790.00	1819.00	1821.00	-----	-----	-----	-----	-----
2055.00	2043.00	3306.00	3304.00	709.00	690.00	-----	-----	-----	-----	-----
1347.00	1325.00	3975.00	3893.00	6528.00	6484.00	-----	-----	-----	-----	-----
4492.00	4449.00	897.00	876.00	1944.00	1934.00	-----	-----	-----	-----	-----
3805.00	3757.00	7648.00	7475.00	3127.00	3138.00	-----	-----	-----	-----	-----
3051.00	3031.00	3635.00	3611.00	16492.00	16238.00	-----	-----	-----	-----	-----
1674.00	1676.00	838.00	824.00	2842.00	2741.00	-----	-----	-----	-----	-----
10030.00	10132.00	3053.00	2987.00	6528.00	6572.00	-----	-----	-----	-----	-----

4492.00	4478.00	10778.00	10756.00	10111.00	9669.00
3278.00	3209.00	1250.00	1234.00	11366.00	11758.00
1347.00	1339.00	2135.00	2108.00	4059.00	3979.00
8691.00	8644.00	5075.00	4993.00	1997.00	1990.00
1347.00	1331.00	897.00	875.00	2570.00	2445.00
4350.00	4341.00	1103.00	1082.00	2570.00	2576.00
12901.00	12601.00	3635.00	3602.00	11366.00	11207.00
3278.00	3339.00	11685.00	11548.00	10111.00	10101.00
12901.00	13224.00	1410.00	1385.00	13982.00	13668.00
3051.00	3072.00	2819.00	2826.00	-----	-----
5487.00	5453.00	19500.00	19348.00	-----	-----
2076.00	2128.00	4140.00	4057.00	-----	-----
1535.00	1554.00	2745.00	2716.00	-----	-----
3051.00	3063.00	1734.00	1710.00	-----	-----
4775.00	4741.00	3053.00	3104.00	-----	-----
10030.00	10021.00	8954.00	8789.00	-----	-----
-----	-----	1147.00	1151.00	-----	-----
-----	-----	2285.00	2234.00	-----	-----
-----	-----	897.00	894.00	-----	-----
-----	-----	688.00	677.00	-----	-----
-----	-----	966.00	955.00	-----	-----
-----	-----	688.00	669.00	-----	-----
FEATURES					
source					
Location/Qualifiers					
1. .169234					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/chromosome="7"					
/map="7p14-15"					
/clone="djs1 (RP11-16G1)"					
/cell_line="Male Blood"					
/clone_lib="RPC-11 Human Male BAC Library"					
392.494					
/rpt_family="MER21"					
complement(1075..1364)					
/rpt_family="Alu"					
complement(2210..2373)					
/standard_name="SWSS2626"					
/note="GenBank Accession Number: G12923"					
2594..2876					
/rpt_family="Alu"					
complement(3252..3702)					
/rpt_family="MLT1"					
complement(4489..4790)					
/rpt_family="Alu"					
complement(5307..5594)					
/rpt_family="Alu"					
5847..6152					
/rpt_family="Alu"					
complement(6970..7082)					
/standard_name="SWSS1000"					
/note="GenBank Accession Number: G00158"					
STS					
repeat_region					
repeat_region					
STS					
repeat_region					
repeat_region					
repeat_region					
repeat_region					
repeat_region					
STS					

```
repeat_region complement(7545..7634)
repeat_region /rpt_family="Alu"
repeat_region complement(8959..9186)
repeat_region /rpt_family="Alu"
repeat_region 10137..10423
repeat_region /rpt_family="Alu"

Query Match 2.6% Score 74; DB 9; Length 169234;
Best Local Similarity 100.0%; Pred. No. 3.7e-27; Mismatches 0; Indels 0; Gaps 0;
Matches 74; Conservative 0;

OY 1982 GGCACCTGTAATCCCGACTTCTGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGAG 2041
|||||
Db 31935 GGCACCTGTAATCCCGACTTCTGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGAG 31876
|||||

OY 2042 GTGAGGTTGCAGT 2055
|||||

Db 31875 GTGAGGTTGCAGT 31862
|||||

RESULT 77
AC093899/C
LOCUS AC093899 172816 bp DNA linear PRI 12-JUN-2002
DEFINITION Homo sapiens BAC clone RP11-724016 from 2, complete sequence.
ACCESSION AC093899 AC068884
VERSION AC093899.3 GI:18497265
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 172816)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
2 (bases 1 to 172816)
Pearman, C., Haakenson, W. and Boyer, E.
The sequence of Homo sapiens BAC clone RP11-724016
Unpublished (2001).
3 (bases 1 to 172816)
Waterston, R.H.
Direct Submission
Submitted (10-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 172816)
Waterston, R.H.
Direct Submission
Submitted (05-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 172816)
Waterston, R.
Direct Submission
Submitted (12-JUN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 5, 2002 this sequence version replaced gi:15625013.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0724016
Drafting Center: WIBR
-----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tatenio, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-724016; actual end is at base position 172816 of RP11-724016.

Unresolved simple sequence repeat from base position 22700 to 25900.

The sequence of AC068884 has been incorporated into AC093899.

FEATURES	Location/Qualifiers
source	1..172816
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-724016"
repeat_region	19..1140
	/clone_lib="RPCI-11"
repeat_region	1141..1437
	/rpt_family="L1"
repeat_region	1438..2657
	/rpt_family="Alu"
repeat_region	2658..2729
	/rpt_family="L1"
repeat_region	2730..3320
	/rpt_family="L1"
repeat_region	3321..3642
	/rpt_family="Alu"
repeat_region	3643..4184
	/rpt_family="L1"
repeat_region	4185..4996
	/rpt_family="ERV1"
repeat_region	5011..5263
	/rpt_family="L1"
repeat_region	5285..5907
	/rpt_family="CRI"
repeat_region	6895..7078
	/rpt_family="L1"
repeat_region	7103..7421
	/rpt_family="Alu"
repeat_region	7807..7952
	/rpt_family="L1"
repeat_region	7981..8289
	/rpt_family="Alu"
repeat_region	8360..8551
	/rpt_family="L1"
repeat_region	8622..8718
	/rpt_family="MIR"

```
repeat_region 8784. .8825 /rpt_family="(TTTTC)n"
repeat_region 8840. .9134 /rpt_family="Alu"
repeat_region 9136. .9174 /rpt_family="Alu"
repeat_region 9861. .9881 /rpt_family="(A)n"
repeat_region 9882. .10166 /rpt_family="Alu"
repeat_region 16502. .16556 /rpt_family="L2"
repeat_region 16886. .17165 /rpt_family="L1"
repeat_region 17179. .17482 /rpt_family="Alu"
repeat_region 18025. .18119 /rpt_family="MIR"
repeat_region 18177. .18471 /rpt_family="Alu"
repeat_region 19312. .19441 /rpt_family="MER1_type"
repeat_region 19731. .20040 /rpt_family="MaLR"
repeat_region 20403. .20472 /rpt_family="L2"
repeat_region 21286. .21344 /rpt_family="(CA)n"
repeat_region 22724. .23977 /rpt_family="(TA)n"
repeat_region 24005. .24171 /rpt_family="(TA)n"
repeat_region 24196. .25007 /rpt_family="(TA)n"
repeat_region 25030. .25205 /rpt_family="(TA)n"
repeat_region 25223. .25398 /rpt_family="(TA)n"
repeat_region 25399. .25455 /rpt_family="AT_rich"
repeat_region 25456. .25919 /rpt_family="(TA)n"
repeat_region 25920. .26206 /rpt_family="Alu"
repeat_region 26893. .26917 /rpt_family="AT_rich"
repeat_region 27194. .27497 /rpt_family="L1"
repeat_region 27498. .27814 /rpt_family="MER2_type"
repeat_region 27815. .28151 /rpt_family="L1"
repeat_region 28152. .28278 /rpt_family="Alu"
repeat_region 28279. .28485 /rpt_family="L1"
repeat_region 28795. .28829 /rpt_family="AT_rich"
repeat_region 28986. .29100 /rpt_family="L2"
repeat_region 29237. .29349 /rpt_family="L1"
repeat_region 29350. .29484 /rpt_family="Alu"
repeat_region 29485. .29589 /rpt_family="L1"
repeat_region 29811. .29874 /rpt_family="L2"
repeat_region 29875. .30284 /rpt_family="L1"
repeat_region 30320. .30522 /rpt_family="L1"
repeat_region 30523. .30847 /rpt_family="L1"

repeat_region /rpt_family="Alu"
30848. .30882 /rpt_family="L1"

Query Match 2.6%; Score 74; DB 9; Length 172816;
Best Local Similarity 100.0%; Pred. NO. 3.7e-27;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTGTAATCCCGAGCTACTTGGAGGCTCGAGGAGAGAAATCGCTTGAACCCAGGAG 2041
|||||
Db 138970 GGCACCTGTAATCCCGAGCTACTTGGAGGCTCGAGGAGAGAAATCGCTTGAACCCAGGAG 138911
|||||

QY 2042 GTGGAGGTTGCAGT 2055
|||||
Db 138910 GTGGAGGTTGCAGT 138897

RESULT 78
AC025518/c
LOCUS 176355 bp DNA linear PRI 29-JUN-2002
DEFINITION Homo sapiens chromosome 17, clone RP11-713H12, complete sequence.
ACCESSION AC025518
VERSION AC025518.8 GI:21629418
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 176355)
Birren,B., Nusbaum,C. and Lander,E.
Unpublished
Homo sapiens chromosome 17, clone RP11-713H12
REFERENCE 2 (bases 1 to 176355)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckigalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczeky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 176355)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouckigalter,B., Brown,A., Camarata,J., Campiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczeky,J., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
```



```
Query Match      2.6%  Score 74;  DB 9;  Length 176355;
Best Local Similarity 100.0%;  Pred. No. 3.7e-27;
Matches 74;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY 1982 GGCACCTATATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTGAACCCAGGAG 2041
DB 152206 GGCACCTATATCCAGCTACTTGGGAGGCTGAGCAGGAGAAATCGCTGAACCCAGGAG 152147
QY 2042 GTGGAGGTTCAGT 2055
DB 152146 GTGGAGGTTCAGT 152133

RESULT 79
AC016772      176932 bp      DNA      linear      PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-567N19 from 2, complete sequence.
DEFINITION AC016772
VERSION AC016772.9 GI:18098289
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 176932)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 176932)
AUTHORS Haglund,K., Abbott,A. and Spalding,L.
TITLE The sequence of Homo sapiens BAC clone RP11-567N19
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 176932)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 176932)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-APR-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 176932)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 6 (bases 1 to 176932)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 7 (bases 1 to 176932)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 9, 2002 this sequence version replaced gi:13677173.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
-----
Summary Statistics
-----
Center project name: H_NH0567N19
-----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-401019; the clone sequenced to the right is AC068884. Actual end of this clone is at base position 176932 of RP11-567N19.

FEATURES

source

```
1..176932
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-567N19"
/clone_lib="RPCI-11"
1..303
/rpt_family="MaLR"
misc_feature 239..639
note="similar to Mus musculus EST AW988274 (NID:g8183383) ug05a08.y1"
misc_feature 305..811
note="match to EST AA769615 (NID:g2820853) ob19g06.s1"
misc_feature 312..507
note="match to EST AA918024 (NID:g3057914) ol73d08.s1"
misc_feature 379..1075
note="similar to Mus musculus EST BF138850 (NID:g10977890)"
misc_feature 404..871
note="similar to Mus musculus EST AA538008 (NID:g2284001) vj34h03.r1"
misc_feature 541..921
note="similar to Mus musculus EST AA106541 (NID:g1657130) ml90d04.r1"
misc_feature 566..1074
note="similar to Mus musculus EST B1455611 (NID:g15246267)"
misc_feature 703..1074
note="match to EST B1550291 (NID:g15437603)"
misc_feature 747..1074
note="similar to Bos taurus EST AV596137 (NID:g9713168)"
misc_feature 763..1076
note="similar to Homo sapiens EST B1459746 (NID:g15250402)"
misc_feature 998..1074
note="similar to EST B1976350 (NID:g16350755)"
1550..1952
/rpt_family="L1"
```

repeat_region	2107..3018	Matches	74: Conservative	0: Mismatches	0: Indels	0: Gaps	0:
repeat_region	/rpt_family="L1"	1982	GGCAGCTGTAATCCAGCTACTTGGAGGCTGAGCAGGAGTAATCGCTTGAACCCAGGAG	2041			
misc_feature	3425..3655						
misc_feature	3714..3942						
misc_feature	/note="match to EST AI801223 (NID:g5366695) to86d01.x1"						
misc_feature	3720..4209						
misc_feature	/note="match to EST AI424984 (NID:g4270915) tg38h07.x1"						
misc_feature	3720..3833						
misc_feature	/note="match to EST AA774547 (NID:g2833881) ai27a12.s1"						
misc_feature	4023..4155						
misc_feature	/note="similar to EST BI976350 (NID:gl6350755)"						
misc_feature	4035..4127						
misc_feature	/note="match to EST BG106656 (NID:g13600502)"						
misc_feature	4038..4127						
misc_feature	/note="match to EST AA774547 (NID:g2833881) ai27a12.s1"						
misc_feature	4089..4101						
misc_feature	/note="similar to Homo sapiens EST BI459746 (NID:g15250402)"						
misc_feature	4274..4360						
misc_feature	/note="match to EST AV648795 (NID:g9869809)"						
misc_feature	4417..4659						
misc_feature	/note="match to EST AV648795 (NID:g9869809)"						
misc_feature	4489..4659						
misc_feature	/note="match to EST AA774547 (NID:g2833881) ai27a12.s1"						
misc_feature	4489..4659						
misc_feature	/note="match to EST BG106656 (NID:g13600502)"						
misc_feature	4489..4659						
misc_feature	/note="match to EST BI550291 (NID:g15437603)"						
misc_feature	4489..4659						
misc_feature	/note="similar to Bos taurus EST AV596137 (NID:g9713168)"						
misc_feature	4489..4659						
misc_feature	/note="similar to EST BI976350 (NID:gl6350755)"						
misc_feature	4489..4659						
misc_feature	/note="similar to Mus musculus EST BI455611 (NID:g15246267)"						
misc_feature	4610..4629						
misc_feature	/note="similar to Homo sapiens EST BI459746 (NID:g15250402)"						
repeat_region	4707..4999						
repeat_region	/rpt_family="Alu"						
repeat_region	4987..5025						
repeat_region	/rpt_family="AT-rich"						
repeat_region	5978..6366						
repeat_region	/rpt_family="ERV1"						
repeat_region	6370..6664						
misc_feature	/rpt_family="Alu"						
misc_feature	6773..6807						
misc_feature	/note="similar to Mus musculus EST BI455611 (NID:g15246267)"						
misc_feature	6773..6806						
misc_feature	/note="match to EST AA774547 (NID:g2833881) ai27a12.s1"						
misc_feature	6773..6806						
misc_feature	/note="similar to Bos taurus EST AV596137 (NID:g9713168)"						
misc_feature	6773..6805						
misc_feature	/note="match to EST AV648795 (NID:g9869809)"						
misc_feature	6773..6805						
misc_feature	/note="match to EST BG106656 (NID:g13600502)"						
misc_feature	6773..6805						
misc_feature	/note="match to EST BI550291 (NID:g15437603)"						
misc_feature	6773..6805						
misc_feature	/note="similar to EST BI976350 (NID:gl6350755)"						
misc_feature	7279..7555						
misc_feature	/note="match to EST BI550291 (NID:g15437603)"						
misc_feature	7279..7454						
misc_feature	/note="similar to EST BI976350 (NID:gl6350755)"						
misc_feature	7313..7555						
misc_feature	/note="match to EST BI552902 (NID:g15440214)"						
misc_feature	7398..7532						
misc_feature	/note="match to EST BG106656 (NID:g13600502)"						
Query Match	2.6%; Score 74; DB 9; Length 176932;						
Best Local Similarity	100.0%; Pred. No. 3.7e-27;						

Submitted (06-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
JOURNAL

REFERENCE
AUTHORS

4 (bases 1 to 182909)
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
 Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
 O'Connor,F., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL

Submitted (14-JUL-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 6, 2002 this sequence version replaced gi:21426269.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WTBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5701

Center clone name: 801_J_18

FEATURES
source

----- Location/Qualifiers
 1..182909
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17"
 /clone="RP11-801J18"
 /clone_lib="RPC1-11 Human Male BAC"
 93..407
 /rpt_family="AluSg"
 complement(1561..1784)
 /rpt_family="L1MB7"
 complement(1789..2051)
 /rpt_family="AluSg"
 complement(4996..5042)
 /rpt_family="L2"
 6912..7047
 /rpt_family="FLAM_C"
 9254..9325
 /rpt_family="L2"
 complement(11585..11631)
 /rpt_family="L2"
 12512..12811
 /rpt_family="AluSg"
 complement(13375..13484)
 /rpt_family="L2"
 complement(13603..13706)
 /rpt_family="MIR"
 complement(14123..14443)
 /rpt_family="AluSx"
 14533..14990
 /rpt_family="L2"
 complement(15253..15339)
 /rpt_family="L2"
 16201..16279
 /rpt_family="MIR"
 complement(18078..18364)
 /rpt_family="AluY"

repeat_region
 complement(18391..18472)
 /rpt_family="L1MB4A"
 18529..18582
 /rpt_family="(TATATG)n"
 complement(19092..19180)
 /rpt_family="L1MB3"
 20169..20453
 /rpt_family="AluJb"
 20454..20497
 /rpt_family="(CA)n"
 20552..20576
 /rpt_family="(CA)n"
 complement(20585..20813)
 /rpt_family="L2"
 complement(20874..21062)
 /rpt_family="L2"
 complement(21389..21395)
 /note="<30 qual SNGL region"
 complement(21443..21745)
 /rpt_family="AluY"
 complement(21508..21516)
 /note="<30 qual SNGL region"
 complement(21530..21538)
 /note="<30 qual SNGL region"
 22759..23068
 /rpt_family="AluSx"
 23122..23128
 /note="<30 qual SNGL region"
 23138..23151
 /note="<30 qual SNGL region"
 complement(23462..23640)
 /rpt_family="L2"
 26691..26909
 /rpt_family="L1MC5"
 complement(27053..27364)
 /rpt_family="AluSg"
 complement(27417..27477)
 /rpt_family="U7"
 complement(27487..27525)
 /rpt_family="L1MA2"
 27526..27586
 /rpt_family="(TA)n"
 complement(27587..27659)
 /rpt_family="L1MA2"
 complement(28085..28239)
 /rpt_family="L1MB3"
 28304..28561
 /rpt_family="AluY"
 28831..28851
 /rpt_family="AT_Rich"
 complement(29182..29789)
 /rpt_family="MER7C"
 30200..30435
 /rpt_family="AluJo"
 30436..30555
 /rpt_family="AluJo/FRAM"
 complement(30603..30912)
 /rpt_family="AluSg"
 complement(32603..32903)
 /rpt_family="AluSx"
 33422..33541
 /rpt_family="MIR"
 complement(35289..35601)
 /rpt_family="AluSp"
 36008..36054
 /rpt_family="L3"
 36097..36121

Query Match

Best Local Similarity 100.0%; Pred. No. 3.7e-27; Length 182909;

Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy

1982 GGCACCTGTATCCAGCTACTTGGGAGGCTGAGCAGGAGATCGCTTGAACCCAGGAG 2041

||||| 88443 GCCACCTGTAATCCAGCTACTTGGAGGCTTGAGCAGGAGAAATCGCTTGAACCCAGGAG 88384
QY 2042 GTGGAGGTTGCAGT 2055
Db 88383 GTGGAGGTTGCAGT 88370

AC116170 186870 bp DNA linear HTG 24-AUG-2002
Homo sapiens chromosome 17 clone RP11-855E10 map 17, WORKING DRAFT
SEQUENCE, 4 unordered pieces.
AC116170
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 186870)
Birren,B., Nussbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone RP11-855E10
Unpublished
2 (bases 1 to 186870)
Birren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhaltier,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Lander,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Melnick,J., Menes,I.,
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 186870)
Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhaltier,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Melnick,J., Menes,I., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 24, 2002 this sequence version replaced gi:19703298.

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26190
Center clone name: 855_E10
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: dye-terminator Big Dye; version 0.960731
Assembly program: Phrap; version 0.960731
Consensus quality: 18544 bases at least Q40
Consensus quality: 185963 bases at least Q30
Consensus quality: 186196 bases at least Q20
Insert size: 188000; agarose-1p
Quality coverage: 17.0 in Q20 bases; agarose-1p
Quality coverage: 17.1 in Q20 bases; sum-of-Contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 16661: contig of 16661 bp in length
* 16662 16761: gap of 100 bp
* 16762 42342: contig of 25581 bp in length
* 42343 42442: gap of 100 bp
* 42443 114057: contig of 71615 bp in length
* 114058 114157: gap of 100 bp
* 114158 186870: contig of 72713 bp in length.
Location/Qualifiers
1..186870
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/chromosome:"17"
/map:"17"
/clone:"RP11-855E10"
/clone.lib:"RPCL-11 Human Male BAC"
1..16661
/note:"assembly_fragment"
clone_end:SP6
vector_side:left
16762..42342
/note:"assembly_fragment"
42443..114057
/note:"assembly_fragment"
114158..186870
/note:"assembly_fragment"
BASE COUNT 51573 a 40568 c 39916 g 54513 t 300 others
ORIGIN
Query Match 2.6% Score 74: DB 2: Length 186870;
Best Local Similarity 100.0%; Pred. No. 3.7e-27;
Matches 74: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
QY 1982 GGACCTGTATCCACGCTACTTGGAGGCTTGAGCAGGAGAAATCGCTTGAACCCAGGAG 2041
Db 138977 GGACCTGTATCCACGCTACTTGGAGGCTTGAGCAGGAGAAATCGCTTGAACCCAGGAG 139016
QY 2042 GTGGAGGTTGCAGT 2055
Db 139037 GTGGAGGTTGCAGT 139050

RESULT 82

DEFINITION
AC092700
AC092700.2 GI:21327455
KEYWORDS
SOURCE
ORGANISM

Homo sapiens chromosome 8, clone RP11-661A3, complete sequence.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 201886)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 8, clone RP11-661A3
Unpublished

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 201886)
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhvalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczký, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (20-JUL-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 201886)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhvalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczký, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced gi:14971446.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 201886)
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhvalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczký, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced gi:14971446.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L11751
Center clone name: 661_A_3

FEATURES
source

Location/Qualifiers
1. 201886
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/chromosome:"8"
/map:"8"
/clone:"RP11-661A3"
/clone_lib:"RPCI-11 Human Male BAC"
4356..4523
/rpt_family:"HAL1b"
complement(4625..4906)
/rpt_family:"AluJo"
8941..9014
/rpt_family:"AluJ/FRAM"
9122..9238
/rpt_family:"(TA)n"
11059..11102
/rpt_family:"L3"
complement(11461..11765)
/rpt_family:"AluSx"
12218..12291
/rpt_family:"tRNA-Glu-GAG"
12584..12885
/rpt_family:"AluSx"
16513..16567
/rpt_family:"L2"
16626..16926
/rpt_family:"AluY"
18349..18494
/rpt_family:"AluSg/x"
18705..19009
/rpt_family:"AluSx"
complement(20040..20229)
/rpt_family:"MIR"
20313..20612
/rpt_family:"AluSx"
20813..21145
/rpt_family:"LIME4A"
22622..22652
/rpt_family:"AT-rich"
22708..22812
/rpt_family:"LIME4A"
24360..24382
/rpt_family:"AT-rich"
24482..24545
/rpt_family:"(TA)n"
complement(24715..25019)
/rpt_family:"AluJb"
25673..25879
/rpt_family:"AluSg/x"
26165..26219
/rpt_family:"(TA)n"
complement(26405..26553)
/rpt_family:"AluJb"
26671..26899
/rpt_family:"MIR"
27947..28028
/rpt_family:"MIR"
30867..31239
/rpt_family:"LIMCC"
31240..31487
/rpt_family:"AluJb"
31488..31862
/rpt_family:"LIMCC"
31923..32003
/rpt_family:"MER21B"
32024..32771
/rpt_family:"MER21B"
32826..33568

TITLE
JOURNAL
COMMENT

Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced gi:14971446.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

```

/rpt_family="LIMCc"
33580. .33721
/rpt_family="MER7A"
repeat_region
complement(33722. .33950)
/rpt_family="MER46A"
repeat_region
33951. .34109
/rpt_family="MER7A"
repeat_region
34110. .34303
/rpt_family="LTRA"
repeat_region
34483. .34764
/rpt_family="LTRA"
repeat_region
34829. .36974
/rpt_family="LIMCc"
repeat_region
37129. .38284
/rpt_family="LIPAG"
repeat_region
38328. .38382
/rpt_family="AT_rich"
repeat_region
38383. .38550
/rpt_family="LIMC4"
repeat_region
38555. .38600
/rpt_family="(CA)n"
repeat_region
38633. .38743
/rpt_family="LIMC4"
repeat_region
39499. .39768
/rpt_family="LIMCc"
repeat_region
39772. .39960
/rpt_family="LIMCc"
repeat_region
complement(40150. .40364)
/rpt_family="L2"
repeat_region
40482. .40510
/rpt_family="AT_rich"
repeat_region
complement(40913. .42148)
/rpt_family="LIPAL5"
repeat_region
complement(42223. .42374)
/rpt_family="AluJb"
repeat_region
complement(42375. .42667)
/rpt_family="AluSg"
repeat_region
complement(42668. .42807)
/rpt_family="AluJb"
repeat_region
complement(43919. .44010)
/rpt_family="L1"
repeat_region
complement(44011. .44322)
/rpt_family="AluSx"
repeat_region
complement(44323. .44921)
/rpt_family="L1"
repeat_region
44923. .45311
/rpt_family="MLT2B1"
repeat_region
45342. .45403
/rpt_family="(TATG)n"
repeat_region
45409. .45463
/rpt_family="MLT2b"
repeat_region
complement(45464. .45891)
/rpt_family="L1"

Query Match 2.6%; Score 74; DB 9; Length 201886;
Best Local Similarity 100.0%; Pred. No. 3.7e-27;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GCACCTGTATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAG 2041
|||||
Db 83334 GCACCTGTATCCAGCTACTTGGAGGCTGAGCAGGAGAAATCGCTTGAACCCAGGAG 83393
|||||

QY 2042 GTGGAGTTGCAGT 2055
|||||
Db 83394 GTGGAGTTGCAGT 83407

RESULT 84
AC024940
LOCUS
DEFINITION Homo sapiens 12 BAC RP11-627K11 (Roswell Park Cancer Institute
Human BAC Library) complete sequence.
ACCESSION AC024940

```

```

VERSION AC024940.39 GI:15809076
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 205952)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Albrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,R.,
Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
Hart,M., Haviak,P., Hawes,A., Hernandez,J., Hernandez,O.,
Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsif,F.,
Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoen,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kurshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisleg,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P.,
Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Ogih,M., Perez,L., Peters,L., Pickens,R., Primmus,E., Pu,L.L.,
Peery,J., Perez,L., Rives,M., Rojars,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savory,G., Scherer,S., Scott,G., Shen,H., Shm.C.,
Shoostari,N., Sisson,I., Sodergren,E., Sonake,T., Sparks,A.,
Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,
Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
Washington,C., Watlington,S., Williams,G., Williamson,A.,
Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.F., Zhou,J.,
Zorrilla,S., Kucheriapati,R., Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 205952)
Worley,K.C.
Direct Submission
Submitted (03-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 205952)
Worley,K.C.
Direct Submission
Submitted (30-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 205952)
Worley,K.C.
Direct Submission
Submitted (02-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 205952)
Worley,K.C.
Direct Submission
Submitted (30-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

```

COMMENT

Baylor Plaza, Houston, TX 77030, USA
On Sep 30, 2001 this sequence version replaced gi:15799479.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES
source

1..205952
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-627K11"
1..2005
/note="overlaps bases 104273..106277 of clone AC068792"
/function="clone overlap"

misc_feature

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region 4039..4125
/rpt_family="AluJo"
repeat_region 4130..4433
/rpt_family="AluSq"
repeat_region 4695..5130
/rpt_family="L2"
repeat_region complement(5131..5432)
/rpt_family="AluSq"
repeat_region 5433..5454
/rpt_family="L2"
repeat_region complement(5455..5748)
/rpt_family="AluSq"
repeat_region 5749..5822
/rpt_family="L2"
repeat_region complement(5833..6112)
/rpt_family="AluJo"
repeat_region 6169..6207
/rpt_family="AT-rich"
repeat_region complement(6416..6702)
/rpt_family="AluX"
repeat_region 7151..7185
/rpt_family="AluY"
repeat_region complement(8477..8768)
/rpt_family="AluJb"
repeat_region complement(8838..8993)
/rpt_family="MERSA"
STS 9177..9471
/standard_name="103578"
repeat_region 9206..9338
/rpt_family="MIR"
repeat_region complement(9422..9506)
/rpt_family="L2"
repeat_region complement(9560..9858)
/rpt_family="AluSg1"
repeat_region complement(10056..10174)
/rpt_family="FLAM_C"
repeat_region 10225..10523
/rpt_family="AluSq"
repeat_region 10568..10850
/rpt_family="AluSq"
repeat_region 10851..10873
/rpt_family="(CAAA)n"
repeat_region 11010..11059
/rpt_family="(TA)n"

Query Match 2.6% Score 74; DB 9; Length 205952;

Best Local Similarity 100.08; Pred. No. 3.7e-27;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1982 GGCACCTCTAATCCAGCTACTTGGAGGCTTGAGCAGGAGATCGCTTGAATCCAGGAG 2041
|||||
Db 33550 GGCACCTCTAATCCAGCTACTTGGAGGCTTGAGCAGGAGATCGCTTGAATCCAGGAG 33609
|||||

QY 2042 GTGGAGGTTGCAGT 2055
|||||

Db 33610 GTGGAGGTTGCAGT 33623
|||||

RESULT 85

CNS01DWD/c

LOCUS

DEFINITION

Human chromosome 14 DNA sequence

from chromosome 14 of Homo sapiens (human), complete sequence.

ACCESSION

AL137128

VERSION

AL137128.4

KEYWORDS

HTG.
human.

228652 bp DNA linear
PK1 40-APR-2001
HAC R-951P22 of library RPI1-11

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
AUTHORS Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
Brottier, P., Cattolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,
Levy, M., Eckenberg, R., Bruls, T., deBerardinis, V., Cruaud, C.,
Gyapay, G., Saurin, W. and Weissbach, J.
Sequencing of the human chromosome 14
Unpublished
2 (bases 1 to 228652)
Genoscope.
Direct Submission
Submitted (30-APR-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On May 2, 2001 this sequence version replaced gi:9755791.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: Segref@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : C-2324M15
Downstream BAC (overlapping the SP6 end) : R-75014 (AC=AL355074)
----- Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 6.64x in Q20 bases; sum-of-contigs

Overall quality chart :
Range : bases
0 :
1 - 9 : 6
10 - 19 : 35
20 - 29 : 86
30 - 39 : 406
40 - 49 : 3708
50 - 59 : 7844
60 - 69 : 11363
70 - 79 : 26580
80 - 89 : 67381
90 - 99 : 111243

Percentage of bases with a quality value >= 40 : 99 %

FEATURES
source
1..228652
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-951P22"
/clone_lib="RPCI-11"
70246..70378
/note="matching EMBL:G07509
RHdb:RH95526
RHdb:RH31991
dbSTS:STS20509
Identified using the e-PCR software (G. Schuler)"
173837..174009
/note="matching EMBL:H49894
RHdb:RH44767
dbSTS:STS37835
Identified using the e-PCR software (G. Schuler)"
175249..175398
/note="matching EMBL:W60094
RHdb:RH76466
dbSTS:STS53553
Identified using the e-PCR software (G. Schuler)"
BASE COUNT 66543 a 47502 c 47687 g 66920 t
ORIGIN

Best Local Similarity 100.0%; Pred. No. 3.7e-27;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1982 GGCACCTCTAATCCAGCTACTTGGGAGGCTGGAGCAGAGAAATCGCTTGAACCCAGGAG 2041
|||||
Db 137366 GGCACCTCTAATCCAGCTACTTGGGAGGCTGGAGCAGAGAAATCGCTTGAACCCAGGAG 137307
|||||
QY 2042 GTGGAGGTTGCAGT 2055
|||||
Db 137306 GTGGAGGTTGCAGT 137293
|||||
RESULT 86
AP004282
LOCUS AP004282 250681 bp DNA linear PRI 04-DEC-2001
DEFINITION Homo sapiens genomic DNA, chromosome 8q23, clone: KB1165G2.
ACCESSION AP004282
VERSION AP004282.2 GI:17298197
KEYWORDS
SOURCE
Homo sapiens pre-pro-B cell cell_line:FLEB 14 - 14 DNA,
clone_lib:Keio BAC library clone:KB1165G2.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1
REFERENCE
AUTHORS Shimizu, N. and Asakawa, S.
TITLE Homo sapiens DNA chromosome 8 SEQUENCE
JOURNAL Published Only in Database (2001)
REFERENCE 2 (bases 1 to 250681)
AUTHORS Shimizu, N. and Asakawa, S.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-2001) Nobuyoshi Shimizu, Keio University, School
of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
160-8582, Japan (E-mail: nshimizu@db.med.keio.ac.jp,
Tel: 81-3-3351-2370, Fax: 81-3-3351-2370)
COMMENT On Dec 3, 2001 this sequence version replaced gi:16303291.
FEATURES
source
1..250681
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8q23"
/clone="KB1165G2"
/cell_line="FLEB 14 - 14"
/cell_type="pre-pro-B cell"
/clone_lib="Keio BAC library"
complement(150..443)
repea_t_region /evidence=not_experimental
/rpt_family="AluSx"
complement(697..880)
repea_t_region /evidence=not_experimental
/rpt_family="MIR"
complement(1245..1561)
repea_t_region /evidence=not_experimental
/rpt_family="AluSg"
1869..2159
repea_t_region /evidence=not_experimental
/rpt_family="AluSx"
complement(2201..2433)
repea_t_region /evidence=not_experimental
/rpt_family="MIR"
2812..2932
repea_t_region /evidence=not_experimental
/rpt_family="MIR"
complement(2928..3017)
repea_t_region /evidence=not_experimental
/rpt_family="L2"
complement(3191..3254)
repea_t_region /evidence=not_experimental
/rpt_family="L2"
3696..3992
repea_t_region /evidence=not_experimental
/rpt_family="AluSg1"

```

repeat_region 4263. .4293
/rpt_family="not_experimental
/rpt_family="(TTTA)n"
complement(4294. .4574)
/evidence=not_experimental
/rpt_family="AluY"
5229. .5498
/evidence=not_experimental
/rpt_family="AluJo"
complement(5615. .5870)
/evidence=not_experimental
/rpt_family="AluJb"
complement(5962. .6152)
/evidence=not_experimental
/rpt_family="MER3"
complement(6266. .6560)
/evidence=not_experimental
/rpt_family="AluJb"
complement(8253. .8442)
/evidence=not_experimental
/rpt_family="MER58A"
complement(8754. .8997)
/evidence=not_experimental
/rpt_family="MER20"
9183. .9210
/evidence=not_experimental
/rpt_family="AT_rich"
9428. .9463
/evidence=not_experimental
/rpt_family="AT_rich"
complement(10944. .11037)
/evidence=not_experimental
/rpt_family="MER58"
complement(11173. .11405)
/evidence=not_experimental
/rpt_family="MER46A"
11785. .11866
/evidence=not_experimental
/rpt_family="L1MC5"
11882. .12020
/evidence=not_experimental
/rpt_family="MER63"
12117. .12151
/evidence=not_experimental
/rpt_family="AT_rich"
12155. .12451
/evidence=not_experimental
/rpt_family="AluJo"
13344. .13572
/evidence=not_experimental
/rpt_family="MIR"
14570. .14633
/evidence=not_experimental
/rpt_family="AT_rich"
complement(15123. .15591)
/evidence=not_experimental
/rpt_family="L1MC3e"
18204. .18500
/evidence=not_experimental
/rpt_family="AluJo"
18840. .18862
/evidence=not_experimental
/rpt_family="AT_rich"
18893. .19050
/evidence=not_experimental
/rpt_family="L1ME2"
19086. .19110
/evidence=not_experimental
/rpt_family="(A)n"
complement(19754. .20039)
/evidence=not_experimental
/rpt_family="AluJb"
22669. .22969

```

```

repeat_region /evidence=not_experimental
/rpt_family="AluSq"
23510. .23619
/evidence=not_experimental
/rpt_family="MIR"
23817. .24117
/evidence=not_experimental
/rpt_family="AluSq"
complement(24121. .24252)
/evidence=not_experimental
/rpt_family="MIR"
complement(24667. .24959)
/evidence=not_experimental
/rpt_family="AluSx"
complement(26086. .26158)
/evidence=not_experimental
/rpt_family="L2"
26183. .26302
/evidence=not_experimental
/rpt_family="MIR"
26303. .26613
/evidence=not_experimental
/rpt_family="AluJb"
26614. .26704
/evidence=not_experimental
/rpt_family="MIR"
26707. .26891
/evidence=not_experimental
/rpt_family="MIR"
complement(27533. .28017)
/evidence=not_experimental
/rpt_family="LTR35"
complement(28078. .28837)
/evidence=not_experimental
/rpt_family="LTR8A"
29031. .29346
/evidence=not_experimental
/rpt_family="AluJo"
29349. .29660
/evidence=not_experimental
/rpt_family="AluJb"
complement(29674. .29730)
/evidence=not_experimental
/rpt_family="L2"
29731. .29780
/evidence=not_experimental
/rpt_family="(CA)n"
complement(29781. .30315)
/evidence=not_experimental
/rpt_family="L2"
complement(31044. .31338)
/evidence=not_experimental
/rpt_family="AluSq"
31497. .31664
/evidence=not_experimental
/rpt_family="HAL1"
31665. .31944
/evidence=not_experimental
/rpt_family="AluSp"
31945. .31968
/evidence=not_experimental
/rpt_family="(CA)n"
31969. .32285
/evidence=not_experimental
/rpt_family="AluY"
32286. .32624
/evidence=not_experimental
/rpt_family="HAL1"

```

Query Match 2.6%; Score 74; DB 9; Length 250681;
 Best Local Similarity 100.0%; Pred.No. 3.7e-27;
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

repeat_region /note="MER3 repeat: matches 18. .137 of consensus"
14816. .15150
repeat_region /note="L2 repeat: matches 879. .1224 of consensus"
15502. .15811
repeat_region /note="Alusg repeat: matches 1. .310 of consensus"
16190. .16382
repeat_region /note="MIR repeat: matches 34. .232 of consensus"
16772. .16872
repeat_region /note="L1M4 repeat: matches 5718. .5824 of consensus"
16906. .17210
repeat_region /note="Alu repeat: matches 1. .304 of consensus"
17465. .18197
repeat_region /note="L2 repeat: matches 1901. .2748 of consensus"
18198. .18327
repeat_region /note="FLAM A repeat: matches 1. .133 of consensus"
18328. .18527
repeat_region /note="L2 repeat: matches 1614. .1901 of consensus"
18722. .19023
repeat_region /note="Alusx repeat: matches 1. .301 of consensus"
20116. .20420
repeat_region /note="Alu repeat: matches 1. .307 of consensus"
20430. .20656
repeat_region /note="Alusg/x repeat: matches 72. .299 of consensus"
21946. .22423
repeat_region /note="Alusx repeat: matches 1. .298 of consensus"
22244. .22373
repeat_region /note="Alusg/x repeat: matches 5. .135 of consensus"
23034. .23339
repeat_region /note="Alusx repeat: matches 7. .312 of consensus"
23059. .23243
misc_feature /note="Sequence from overlapping clone RP11-533020
(AL513008). Assembly confirmed by restriction digest."
23708. .24046
misc_feature /note="match: GSS: Em:A0770255"
23760. .24041
repeat_region /note="Alusx repeat: matches 1. .291 of consensus"
24548. .24864
repeat_region /note="Alusg repeat: matches 1. .313 of consensus"
25207. .25358
repeat_region /note="L1MC5 repeat: matches 7779. .7929 of consensus"
25359. .25392
repeat_region /note="17 copies 2 mer gt 94% conserved"
25394. .25537
repeat_region /note="Alusg repeat: matches 143. .284 of consensus"
25538. .25856
repeat_region /note="Alu/b8 repeat: matches 1. .306 of consensus"
25857. .25993
repeat_region /note="Alusg repeat: matches 1. .143 of consensus"
26082. .26209
repeat_region /note="L1MD3 repeat: matches 7612. .7740 of consensus"
26244. .26526
repeat_region /note="L1MD2 repeat: matches 5907. .6190 of consensus"
26527. .26826
repeat_region /note="Alusx repeat: matches 1. .294 of consensus"
26827. .26988
repeat_region /note="L1MD2 repeat: matches 5746. .5907 of consensus"
27450. .27589
repeat_region /note="L1P4 repeat: matches 5677. .5821 of consensus"
27897. .28025
repeat_region /note="L2 repeat: matches 2383. .2522 of consensus"
28089. .28389
repeat_region /note="Alu repeat: matches 2. .302 of consensus"
28485. .28537
repeat_region /note="L2 repeat: matches 2648. .2698 of consensus"
28493. .28610

Query Match 2.6%; Score 73; DB 9; Length 79319;
Best Local Similarity 100.0%; Pred. No. 1.2e-26;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1982 GCCACCTGTATCCAGCTACTTGGGAGGCTGAGCGAGGAGATCGCTTGAACCCAGGAG 2041
Db 70901 GCCACCTGTATCCAGCTACTTGGGAGGCTGAGCGAGGAGATCGCTTGAACCCAGGAG 70960

Qy 2042 GTGGAGGTTGCAG 2054
Db 70961 GTGGAGGTTGCAG 70973
RESULT 88
AC009702
LOCUS AC009702 Homo sapiens, clone RP11-19711, complete sequence.
DEFINITION AC009702
ACCESSION AC009702
VERSION AC009702.12 GI:20901910
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 96165)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-19711
Unpublished
2 (bases 1 to 96165)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kani,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychalek,J., J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 96165)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boquslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamaizais,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 96165)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boquslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,

Ginde, S., Gord, S., Govette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamata, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., Murphy, K., McKernan, K., Meldrum, J., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rossetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.	Direct Submission	
	Submitted (17-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
	On May 17, 2002 this sequence version replaced gi:20128778.	
	All repeats were identified using RepeatMasker:	
	Smith, A.F.A. & Green, P. (1996-1997)	
	http://ftp.genome.washington.edu/RM/RepeatMasker.html	
TITLE		
JOURNAL		
COMMENT		

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L2348
Center clone name: 197_I_11

Only the last 96.2 kilobases of this clone are being submitted. The remainder overlaps accession number AC021915 [WICGR project L4214].

[illegible]

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Submitted (09-AUG-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
6 (bases 1 to 144355)
Waterston,R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 24, 2001 this sequence version replaced gi:13270791.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0475H17

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Oseogawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Cataneese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(<http://bacpac.med.buffalo.edu>)
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-471D6; the clone sequenced
to the right is RP11-479L11, 200 bp overlap. Actual end of this
clone is at base position 49393 of RP11-479L11.

FEATURES

source	1..144355	/organism="Homo sapiens"
	/db_xref="taxon:9606"	
	/chromosome="2"	
	/map="2"	
	/clone="RP11-475H17"	
	/clone_lib="RPCI-11"	
repeat_region	405..445	/rpt_family="AT_rich"
repeat_region	1826..2062	/rpt_family="ERVL"
repeat_region	3399..3600	/rpt_family="MaLR"
repeat_region	4181..4244	/rpt_family="MIR"
repeat_region	5716..5913	/rpt_family="MIR"
repeat_region	5983..6010	/rpt_family="AT_rich"
repeat_region	6039..6272	/rpt_family="L1"
repeat_region	6476..7408	/rpt_family="ERVI"

repeat_region	12965..13128	/rpt_family="L1"
repeat_region	13129..13445	/rpt_family="Alu"
repeat_region	13446..13999	/rpt_family="L1"
repeat_region	14000..14386	/rpt_family="L1"
repeat_region	14387..14667	/rpt_family="Alu"
repeat_region	14668..14723	/rpt_family="L1"
repeat_region	14724..15471	/rpt_family="L1"
repeat_region	15473..15608	/rpt_family="L1"
repeat_region	15639..18141	/rpt_family="L1"
repeat_region	18142..18489	/rpt_family="MaLR"
repeat_region	18642..18662	/rpt_family="AT_rich"
repeat_region	20137..20494	/rpt_family="MaLR"
repeat_region	20500..20566	/rpt_family="AT_rich"
repeat_region	21874..22216	/rpt_family="MERL_type"
repeat_region	22478..22536	/rpt_family="L2"
repeat_region	22712..23154	/rpt_family="ERV1"
repeat_region	23618..23902	/rpt_family="L1"
repeat_region	24035..24055	/rpt_family="AT_rich"
repeat_region	24243..24487	/rpt_family="MIR"
repeat_region	24807..25180	/rpt_family="MaLR"
repeat_region	25519..25816	/rpt_family="Alu"
repeat_region	25904..25966	/rpt_family="AT_rich"
repeat_region	26064..26379	/rpt_family="Alu"
repeat_region	32986..33157	/rpt_family="ERVL"
repeat_region	33404..33690	/rpt_family="L1"
repeat_region	35170..36476	/rpt_family="L1"
repeat_region	37329..37359	/rpt_family="AT_rich"
repeat_region	37537..37566	/rpt_family="AT_rich"
repeat_region	37756..38061	/rpt_family="Alu"
repeat_region	38264..38503	/rpt_family="ERV1"
repeat_region	38545..38585	/rpt_family="L2"
repeat_region	39025..39244	/rpt_family="ERV1"
repeat_region	39413..39589	/rpt_family="ERV1"
repeat_region	39782..39946	/rpt_family="MIR"
repeat_region	40393..40665	/rpt_family="Alu"
repeat_region	41151..41855	/rpt_family="ERV1"

```

repeat_region 41945..42232
                /rpt_family="Alu"
repeat_region 42828..43210
                /rpt_family="ERVL"
repeat_region 43211..43510
                /rpt_family="L1"
repeat_region 43511..43812
                /rpt_family="ERVL"
repeat_region 44561..44632
                /rpt_family="L2"
repeat_region 44633..45357
                /rpt_family="ERVL"
repeat_region 45358..46188
                /rpt_family="L2"
repeat_region 46279..46514
                /rpt_family="Alu"
repeat_region 46874..46958
                /rpt_family="MaLR"
repeat_region 47176..48678
                /rpt_family="L1"

```

Query Match 2.6% Score 73; DB 9; Length 144355;
 Best Local Similarity 100.0%; Pred. No. 1.2e-26;
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1993 TCCAGCTACTTGGAGGCTGAGCGAGAGAGTAATCGCTTGAACCCAGGAGGTGGAGTTGC 2052
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 135951 TCCAGCTACTTGGAGGCTGAGCGAGAGAGTAATCGCTTGAACCCAGGAGGTGGAGTTGC 136010
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 2053 AGTAAGCTGAGAT 2065
      |||||||||||||||
DB 136011 AGTAAGCTGAGAT 136023

```

RESULT 91

```

AC023536
LOCUS AC023536 155378 bp DNA linear HTG 09-MAY-2001
DEFINITION Homo sapiens chromosome 8 clone RP11-219D23 map 8, WORKING DRAFT
SEQUENCE, 8 unordered pieces.
AC023536
VERSION AC023536.5 GI:13958545
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
1 (bases 1 to 155378)
Homo sapiens chromosome 8, clone RP11-219D23
Unpublished
2 (bases 1 to 155378)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bedalov,F., Boguski,M.,
Bouhassira,D., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Chen,Y., Colangelo,M., Collins,S., Collins,S., Cooke,P.,
DeRubeis,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenech,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardy,J., Glendon,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hago,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karas,A.,
Klein,J., Lander,E., Leach,K., Leach,K., Levine,R.,
Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N., McCarthy,M.,
McEwen,P., McGurk,A., McKernan,K., McPherson,R., Meldrum,J.,
Menees,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olliver,T.M.,
Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
TITLE

```

JOURNAL

COMMENT

Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 6, 2001 this sequence version replaced gi:1184166.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIDR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L5446
 Center clone name: 219.D.23

----- Summary Statistics
 Sequencing vector: M13; M7815; 2% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 153109 bases at least Q40
 Consensus quality: 154194 bases at least Q30
 Consensus quality: 154534 bases at least Q20
 Insert size: 157000; agarose-fp
 Insert size: 154678; sum-of-contigs
 Quality coverage: 11.3 in Q20 bases; agarose-fp
 Quality coverage: 11.4 in Q20.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 40: contig of 40 bp in length
 * 41 140: gap of 100 bp
 * 141 1311: contig of 1171 bp in length
 * 1312 1411: gap of 100 bp
 * 1412 3166: contig of 1755 bp in length
 * 3167 3266: gap of 100 bp
 * 3267 59470: contig of 56204 bp in length
 * 59471 59570: gap of 100 bp
 * 59571 79596: contig of 20026 bp in length
 * 79597 79696: gap of 100 bp
 * 79697 109382: contig of 29686 bp in length
 * 109383 109482: gap of 100 bp
 * 109483 145165: contig of 35683 bp in length
 * 145166 145265: gap of 100 bp
 * 145266 155378: contig of 10113 bp in length.

FEATURES

```

source
1..155378
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-219D23"
/clone_lib="RP11 Human Male BAC"
misc_feature
1..40
/note="assembly_fragment"
clone_end:SP6
vector_side:left
141..1311
/note="assembly_fragment"
1412..3166
/note="assembly_fragment"
3267..59470
/note="assembly_fragment"
59571..79596
/note="assembly_fragment"
79697..109382
/note="assembly_fragment"
109483..145165
/note="assembly_fragment"

```



```
misc_feature 145266..155378
              /note="assembly_fragment
              clone_end:T7
              vector_side:right"
BASE COUNT 46893 a 30633 c 30876 g 46276 t 700 others
ORIGIN

Query Match      2.6%  Score 73;  DB 2;  Length 155378;
Best Local Similarity 100.0%;  Pred. No. 1.2e-26;
Matches 73;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY 1983 GCACCTGTAAATCCAGCTACTTGGGAGGCTGAGCAGGAGATGCGTTGAACCCAGGAGG 2042
      |||
DB 116692 GCACCTGTAAATCCAGCTACTTGGGAGGCTGAGCAGGAGATGCGTTGAACCCAGGAGG 116751
      |||

QY 2043 TGGAGGTTGCAGT 2055
      |||
DB 116752 TGGAGGTTGCAGT 116764

RESULT 92
AC009305
LOCUS      167116 bp      DNA      linear      PRI 07-OCT-2000
DEFINITION Homo sapiens BAC clone RP11-153B21 from 2, complete sequence.
ACCESSION AC009305
VERSION   AC009305.2 GI:9857572
KEYWORDS HTG.
SOURCE    Homo sapiens.
  ORGANISM Homo sapiens.
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
  AUTHORS Sulston,J.E. and Waterston,R.
  TITLE   Toward a complete human genome sequence
  JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE 99063792
  PUBMED  9847074

REFERENCE 2 (bases 1 to 167116)
AUTHORS  Kyung,K., Maupin,R., Elliot,G. and Long,N.
TITLE    The sequence of Homo sapiens BAC clone RP11-153B21
JOURNAL  Unpublished
REFERENCE 3 (bases 1 to 167116)
AUTHORS  Waterston,R.H.
TITLE    Direct Submission
JOURNAL  Submitted (13-AUG-1999) Genome Sequencing Center, Washington
        University School of Medicine, 4444 Forest Park Parkway, St. Louis,
        MO 63108, USA
REFERENCE 4 (bases 1 to 167116)
AUTHORS  Waterston,R.H.
TITLE    Direct Submission
JOURNAL  Submitted (19-AUG-2000) Genome Sequencing Center, Washington
        University School of Medicine, 4444 Forest Park Parkway, St. Louis,
        MO 63108, USA
REFERENCE 5 (bases 1 to 167116)
AUTHORS  Waterston,R.H.
TITLE    Direct Submission
JOURNAL  Submitted (10-SEP-2000) Genome Sequencing Center, Washington
        University School of Medicine, 4444 Forest Park Parkway, St. Louis,
        MO 63108, USA
REFERENCE 6 (bases 1 to 167116)
AUTHORS  Waterston,R.H.
TITLE    Direct Submission
JOURNAL  Submitted (07-OCT-2000) Department of Genetics, Washington
        University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
        On Aug 19, 2000 this sequence version replaced gi:5732166.
COMMENT  ----- Genome Center
        Center: Washington University Genome Sequencing Center
        Center code: WUGSC
        Web site: http://genome.wustl.edu/gsc
        Contact: saplens@watson.wustl.edu
        ----- Summary Statistics
        -----
        Center project name: H_NH0153B21
        -----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenoh,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-541A15; the clone sequenced to the right is RP11-62F14. Actual start of this clone is at base position 1 of RP11-153B21; actual end is at base position 167116 of RP11-153B21.

FEATURES	Location/Qualifiers
source	1..167116
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-153B21"
	/clone_lib="RPCI-11"
repeat_region	1483..2196
	/rpt_family="Retroviral"
repeat_region	2460..2480
	/rpt_family="(CAAAA)n"
repeat_region	3940..4032
	/rpt_family="(CA)n"
repeat_region	5461..5585
	/rpt_family="MIR"
repeat_region	5786..5815
	/rpt_family="AT-rich"
repeat_region	7025..7137
	/rpt_family="L1"
repeat_region	7178..7486
	/rpt_family="MERL_type"
repeat_region	7488..7715
	/rpt_family="L1"
repeat_region	8235..8426
	/rpt_family="MIR"
repeat_region	9134..9227
	/rpt_family="L1"
repeat_region	9924..9981
	/rpt_family="MIR"
repeat_region	10061..10155
	/rpt_family="Alu"
repeat_region	10222..10282
	/rpt_family="MIR"
repeat_region	11425..11565
	/rpt_family="MERL_type"
repeat_region	11586..11866
	/rpt_family="Alu"

repeat_region	11867..11991	/rpt_family="MER1_type"
repeat_region	12108..12263	/rpt_family="MER1_type"
repeat_region	13912..14201	/rpt_family="Alu"
repeat_region	14588..14891	/rpt_family="Alu"
repeat_region	14893..15914	/rpt_family="L1"
repeat_region	16248..16622	/rpt_family="L1"
repeat_region	16654..16969	/rpt_family="MaLR"
repeat_region	17028..17219	/rpt_family="L1"
repeat_region	17705..18003	/rpt_family="Alu"
repeat_region	18232..18533	/rpt_family="L1"
repeat_region	18740..18966	/rpt_family="Alu"
repeat_region	19495..19801	/rpt_family="Alu"
repeat_region	19813..19928	/rpt_family="Other"
repeat_region	20009..20181	/rpt_family="L1"
repeat_region	20554..21109	/rpt_family="LTR19B"
repeat_region	23356..23414	/rpt_family="L2"
repeat_region	25792..25889	/rpt_family="MIR"
misc_feature	27584..28069	/note="match to EST AA426194 (NID:g2107534) zwlf02.rl"
repeat_region	27700..27828	/rpt_family="L1"
misc_feature	27814..28226	/note="match to EST AI867228 (NID:g5540244) wa01f10.xl"
repeat_region	28785..29014	/rpt_family="L1"
repeat_region	29015..29301	/rpt_family="Alu"
repeat_region	29302..30552	/rpt_family="L1"
repeat_region	30587..30882	/rpt_family="Other"
repeat_region	31514..31907	/rpt_family="L1"
repeat_region	31982..32649	/rpt_family="L1"
repeat_region	32650..33025	/rpt_family="MaLR"
repeat_region	33026..33116	/rpt_family="L1"
repeat_region	33268..33291	/rpt_family="AT-rich"
repeat_region	33382..33635	/rpt_family="L1"
repeat_region	33683..34038	/rpt_family="MER2_type"
repeat_region	34097..34410	/rpt_family="L1"
repeat_region	34411..34613	/rpt_family="MER2_type"
repeat_region	34614..35269	/rpt_family="L1"
repeat_region	35402..35564	/rpt_family="L1"
repeat_region	35714..36147	/rpt_family="MER2_type"
repeat_region	37176..37360	
repeat_region	/rpt_family="L2" 37939..37974 /rpt_family="(TGG)n"	
Query Match	2.6%; Score 73; DB 9; Length 167116;	
Best Local Similarity	100.0%; Pred. No. 1.2e-26;	
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1983	GCACCTGTAATCCAGCTACTTGGAGGCTCAGGAGCAAGTAATCGCTTGAAACCCAGGAGG 2042
Db	88394	GCACCTGTAATCCAGCTACTTGGAGGCTCAGGAGCAAGTAATCGCTTGAAACCCAGGAGG 88453
Qy	2043	TGGAGGTGCAGT 2055
Db	88454	TGGAGGTGCAGT 88466
RESULT 93		
AC079360	176176 bp DNA linear HTG 31-JUL-2002	
LOCUS	Homo sapiens chromosome 12 clone RP11-7M8, WORKING DRAFT SEQUENCE,	
DEFINITION	12 unordered pieces.	
ACCESSION	AC079360	
VERSION	AC079360.22 GI:22002362	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota: Metazoa; Chordata; Craniata: Vertebrata: Euteleostomi;	
AUTHORS	Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.	
	1 (bases 1 to 176176)	
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,	
	Aisbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,	
	Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,	
	Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,	
	Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,	
	Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,	
	Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,	
	Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,	
	Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,	
	Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinoh,H.H.,	
	Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,	
	Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escott,M.,	
	Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frank,P.,	
	Gabis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,	
	Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,	
	Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,	
	Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,	
	Homsif., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,B.E.,	
	Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,	
	Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,	
	Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,	
	Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lounsqed,H.,	
	Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,	
	Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,	
	Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,	
	Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,	
	Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,	
	Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwona,G.,	
	Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,	
	Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ron,Y.,	
	Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savory,G.,	
	Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,	
	Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,	
	Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,	
	Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,	
	Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,K., Wang,Q.,	
	Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,	
	Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,	
	Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,	
	Weinstock,G. and Gibbs,R.	
	Direct Submission	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 176176)	

AUTHORS
TITLE
JOURNAL

Worley, K.C.
Direct Submission
Submitted (29-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 1761176)

REFERENCE
AUTHORS
TITLE
JOURNAL

Worley, K.C.
Direct Submission
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 29, 2002 this sequence version replaced gi:19718706.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Drafting Center Code: BCM
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCCP
Center clone name: RP11-7M8
----- Summary Statistics
Sequencing vector: Plasmid;
Sequencing vector: M13;
Chemistry: Dye-Primer Bodipy: 18% of reads
Chemistry: Dye-Terminator Big Dye: 82% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 169109 bases at least Q40
Consensus quality: 171607 bases at least Q30
Consensus quality: 173502 bases at least Q20
Estimated insert size: 163861; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2163: contig of 2163 bp in length
* 2164 2263: gap of unknown length
* 2264 4836: contig of 2573 bp in length
* 4837 4936: gap of unknown length
* 4937 7285: contig of 2349 bp in length
* 7286 7385: gap of unknown length
* 7386 9800: contig of 2415 bp in length
* 9801 9901: gap of unknown length
* 9901 14017: contig of 4117 bp in length
* 14018 14117: gap of unknown length
* 14118 19883: contig of 5766 bp in length
* 19884 19984: gap of unknown length
* 19984 25707: contig of 5723 bp in length
* 25707 25806: gap of unknown length
* 25807 33908: contig of 8102 bp in length
* 33909 34008: gap of unknown length
* 34009 42136: contig of 8128 bp in length
* 42137 42237: gap of unknown length
* 42237 59945: contig of 17709 bp in length
* 59946 60045: gap of unknown length
* 60046 82634: contig of 22589 bp in length
* 82635 82734: gap of unknown length
* 82735 176176: contig of 93442 bp in length.

FEATURES
source

Location/Qualifiers
1..176176
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-7M8"

BASE COUNT

40817 a 46113 c 45951 g 42181 t 1114 others

ORIGIN

Query Match 2.6%; Score 73; DB 2; Length 176176;
Best Local Similarity 100.0%; Pred. No. 1.2e-25;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1983 GCACCTGTAATCCCGAGCTACTTGGGAGCTGAGGAGGAGGAGTGCCTTGAACCCAGGAGG 2042
|||||
Db 60822 GCACCTGTAATCCCGAGCTACTTGGGAGCTGAGGAGGAGGAGTGCCTTGAACCCAGGAGG 60881
|||||
QY 2043 TGGAGGTTGCAGT 2055
|||||
Db 60882 TGGAGGTTGCAGT 60894
|||||

RESULT 94
AL162211/c
LOCUS
DEFINITION Homo sapiens chromosome 13 clone RP11-157H17, *** SEQUENCING IN
PROGRESS ***, 16 unordered pieces.
ACCESSION AL162211
VERSION AL162211.5 GI:9863623
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 179470)
Direct Submission
Burton, J.
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:9212905.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA157H17
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye: 100% of reads
Consensus quality: 170263 bases at least Q40
Consensus quality: 173209 bases at least Q30
Consensus quality: 175040 bases at least Q20
Insert size: 177970; sum-of-contigs
Quality coverage: 3.74x in Q20 bases; agarose-fp
coverage: 3.12x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 20111: contig of 20111 bp in length
* 20112 20211: gap of 100 bp
* 20212 22684: contig of 2473 bp in length
* 22685 22784: gap of 100 bp
* 22785 26050: contig of 3266 bp in length
* 26051 26150: gap of 100 bp
* 26151 50870: contig of 24720 bp in length
* 50871 50970: gap of 100 bp
* 50971 56714: contig of 5744 bp in length
* 56715 56814: gap of 100 bp
* 56815 79207: contig of 22392 bp in length
* 79207 79306: gap of 100 bp

* 79307 82214: contig of 2908 bp in length
 * 82215 82314: gap of 100 bp
 * 86916: contig of 4602 bp in length
 * 86917 87016: gap of 100 bp
 * 87017 98674: contig of 11658 bp in length
 * 98675 98774: gap of 100 bp
 * 98775 112648: contig of 13874 bp in length
 * 112649 112748: gap of 100 bp
 * 112749 126894: contig of 14146 bp in length
 * 126895 126994: gap of 100 bp
 * 126995 140436: contig of 13442 bp in length
 * 140437 140536: gap of 100 bp
 * 140537 148268: contig of 7732 bp in length
 * 148269 148368: gap of 100 bp
 * 148369 151247: contig of 2879 bp in length
 * 151248 151347: gap of 100 bp
 * 151348 160829: contig of 9482 bp in length
 * 160830 160929: gap of 100 bp
 * 160930 179470: contig of 18541 bp in length.

FEATURES

source

Location/Qualifiers
 1. .179470
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone_lib="RPC1-11.1"
 1. .20111
 /note="assembly_fragment:01022
 fragment_chain:1"
 20212. .22684
 /note="assembly_fragment:00962
 fragment_chain:1"
 22785. .26050
 /note="assembly_fragment:00754
 fragment_chain:1"
 26151. .50870
 /note="assembly_fragment:00208
 fragment_chain:2"
 50971. .56714
 /note="assembly_fragment:01315
 fragment_chain:2"
 56815. .79206
 /note="assembly_fragment:00254
 fragment_chain:3"
 79307. .82214
 /note="assembly_fragment:00835
 fragment_chain:3"
 82315. .86916
 /note="assembly_fragment:00293
 fragment_chain:4"
 87017. .98674
 /note="assembly_fragment:00855
 fragment_chain:4"
 98775. .112648
 /note="assembly_fragment:01017
 fragment_chain:5"
 112749. .126894
 /note="assembly_fragment:01010
 fragment_chain:5"
 126895. .140436
 /note="assembly_fragment:01132
 fragment_chain:6"
 140537. .148268
 /note="assembly_fragment:00107
 fragment_chain:6"
 148369. .151247
 /note="assembly_fragment:00075"
 151348. .160829
 /note="assembly_fragment:00269"
 160930. .179470
 /note="assembly_fragment:00989"

BASE COUNT 54009 a 35515 c 34477 g 53955 t 1514 others
 ORIGIN

Query Match 2.6%: Score 73; DB 2: Length 179470;
 Best Local Similarity 100.08; Pred. No. 1.2e-26;
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1983 GCACCTGTAATCCCGAGCTACTTGGAGGCTGAGCGAGGAGGAGGAGG 2042
 |||||
 Db 95945 GCACCTGTAATCCCGAGCTACTTGGAGGCTGAGCGAGGAGGAGGAGG 95886
 |||||

QY 2043 TGGAGTTGCCAGT 2055
 |||||
 Db 95885 TGGAGTTGCCAGT 95873

RESULT 95
 AC021915/c
 LOCUS 183101 bp DNA linear PRI 40-APR-2002
 DEFINITION Homo sapiens chromosome 8, clone RP11-26M5, complete sequence.
 ACCESSION AC021915
 VERSION AC021915.11 GI:20128647
 KEYWORDS HTG
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 183101)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 8, clone RP11-26M5
 Unpublished
 2 (bases 1 to 183101)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 Dearellano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
 Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karas,A., Klein,J.,
 Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
 McPheeters,R., Meldrim,J., Menus,L., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 183101)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
 Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
 Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karas,A., Kells,C., LaRoque,K., Lamazatus,R.,
 Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,K.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Riese,C., Kirov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,K.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1993 GCACCTGTATCCAGCTACTTGGAGGCTGAGCAGAGAAATCGCTTGAACCCAGGAGG 2042
|||||
Db 11407 GCACCTGTATCCAGCTACTTGGAGGCTGAGCAGAGAAATCGCTTGAACCCAGGAGG 11348
|||||

QY 2043 TGGAGGTTGCGAGT 2055
|||||

Db 11347 TGGAGGTTGCGAGT 11335
|||||

RESULT 96
AL354668/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AL354668
Human DNA sequence from clone RP11-506P24 on chromosome 13,
complete sequence.
AL354668
AL354668.13 GI:12717990
HTG.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 191652)
Kay,M.
Direct Submission
Submitted (21-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Feb 8, 2001 this sequence version replaced gi:12580982.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-506P24 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3 6
This sequence is the entire insert of clone RP11-506P24 The true
left end of clone RP11-57H17 is at 67628 in this sequence.
FEATURES
source
Location/Qualifiers
1..191652
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-506P24"
/clone_lib="RPCI-11.2"
562..1027
/note="L2 repeat: matches 1901..2407 of consensus"
repeat_region
1778..2101
/note="AluX repeat: matches 1..311 of consensus"
repeat_region
2217..2290
/note="MER41B repeat: matches 175..547 of consensus"
repeat_region
5887..5930
/note="L2 repeat: matches 2572..2710 of consensus"
44416..44495

repeat_region
6161..8565
/note="L1PA16 repeat: matches 3670..6139 of consensus"
10594..11173
/note="MT2CB repeat: matches 1..502 of consensus"
11174..11246
/note="MIR repeat: matches 25..99 of consensus"
11947..12156
/note="L1PA15 repeat: matches 5940..6153 of consensus"
14745..14978
/note="MIR repeat: matches 32..262 of consensus"
15709..15778
/note="L2 repeat: matches 2679..2750 of consensus"
16231..16413
/note="MT1H repeat: matches 110..288 of consensus"
16445..16557
/note="MT1H repeat: matches 382..495 of consensus"
18920..19339
/note="MSTB repeat: matches 4..421 of consensus"
20087..20378
/note="AluX repeat: matches 1..295 of consensus"
20679..21301
/note="L1MC3 repeat: matches 7114..7740 of consensus"
21353..22028
/note="L1MA7 repeat: matches 5602..6279 of consensus"
22548..23928
/note="PTR5 repeat: matches 743..2438 of consensus"
25784..26060
/note="HERVL repeat: matches 4300..4592 of consensus"
26611..26873
/note="HERVL repeat: matches 3367..3648 of consensus"
26926..27092
/note="HERVL repeat: matches 3100..3269 of consensus"
27143..27443
/note="MERVL repeat: matches 2736..3042 of consensus"
27972..28082
/note="HERVL repeat: matches 2158..2270 of consensus"
29273..29761
/note="HAL1 repeat: matches 1145..1607 of consensus"
29918..30090
/note="MIR repeat: matches 20..196 of consensus"
3027..32082
/note="28 copies 2 mer aa 78% conserved"
32377..32558
/note="MIR repeat: matches 10..214 of consensus"
32658..33379
/note="L1MC1 repeat: matches 5592..6332 of consensus"
33722..34012
/note="AluX repeat: matches 2..293 of consensus"
34152..34243
/note="MIR repeat: matches 28..131 of consensus"
34225..34334
/note="MER5B repeat: matches 68..185 of consensus"
34349..34530
/note="MER5A repeat: matches 1..188 of consensus"
34815..34879
/note="L2 repeat: matches 2641..2703 of consensus"
34897..35200
/note="76 copies 4 mer tata 80% conserved"
36514..36545
/note="8 copies 4 mer caca 87% conserved"
36746..36842
/note="MER40A repeat: matches 1..97 of consensus"
38934..39176
/note="AluSg/x repeat: matches 168..296 of consensus"
41131..41428
/note="AluSg repeat: matches 1..298 of consensus"
42548..42734
/note="MIR repeat: matches 65..262 of consensus"
43086..43201
/note="29 copies 4 mer aagg 84% conserved"
44272..44415
/note="L2 repeat: matches 2572..2710 of consensus"
44416..44495

repeat_region /note="MADE1 repeat: matches 1. .80 of consensus"
44496. .44931
repeat_region /note="L2 repeat: matches 2112. .2572 of consensus"
45513. .47217
repeat_region /note="L1PA6 repeat: matches 4434. .6143 of consensus"
47218. .47396
repeat_region /note="L1PB3 repeat: matches 5956. .6126 of consensus"
50159. .51245
repeat_region /note="L2 repeat: matches 1609. .2750 of consensus"
51426. .51726
repeat_region /note="L2 repeat: matches 1199. .1530 of consensus"
51733. .51845
repeat_region /note="L2 repeat: matches 2. .118 of consensus"
51773. .52349
repeat_region /note="MER39b repeat: matches 4. .577 of consensus"
52355. .52386
repeat_region /note="16 copies 2 mer aa 90% conserved"
52499. .52801
repeat_region /note="AluSp repeat: matches 1. .313 of consensus"
53175. .53430
repeat_region /note="L2 repeat: matches 1. .256 of consensus"
53486. .53621
repeat_region /note="L2 repeat: matches 1617. .1753 of consensus"
55490. .56705
repeat_region /note="L1MA5 repeat: matches 5064. .6300 of consensus"
56706. .57003
repeat_region /note="AluSq repeat: matches 1. .304 of consensus"
57004. .57087
repeat_region /note="L1MA5 repeat: matches 4980. .5064 of consensus"
57088. .57374
repeat_region /note="AluJo repeat: matches 1. .284 of consensus"
57375. .58079
repeat_region /note="L1MA5 repeat: matches 4339. .4980 of consensus"
58080. .58374
repeat_region /note="AluSk repeat: matches 1. .294 of consensus"
58375. .58458
repeat_region /note="L1MA5 repeat: matches 4255. .4339 of consensus"
58489. .59221
repeat_region /note="L1PA12 repeat: matches 5411. .6153 of consensus"
59217. .60773
repeat_region /note="L1MA5 repeat: matches 2675. .4268 of consensus"
60988. .61059
repeat_region /note="L1M2 repeat: matches 354. .423 of consensus"
61512. .62087
repeat_region /note="L1M2 repeat: matches -668. .-100 of consensus"
62833. .63254
repeat_region /note="MSTC repeat: matches 3. .405 of consensus"
64230. .64526
repeat_region /note="AluSg repeat: matches 1. .296 of consensus"
66309. .66499
repeat_region /note="L1M2 repeat: matches 6154. .6340 of consensus"
67079. .67376
repeat_region /note="AluSk repeat: matches 2. .304 of consensus"
67502. .67541
repeat_region /note="10 copies 4 mer atat 82% conserved"
67822. .67838
repeat_region /note="L2 repeat: matches 2732. .2749 of consensus"
67839. .68126
repeat_region /note="AluJo repeat: matches 14. .300 of consensus"
68127. .68257
repeat_region /note="L2 repeat: matches 2616. .2732 of consensus"
70069. .70100
repeat_region /note="16 copies 2 mer gt 84% conserved"
70410. .73012
repeat_region /note="L1M1 repeat: matches 3067. .5737 of consensus"
73878. .73905
repeat_region /note="14 copies 2 mer aa 89% conserved"

Query Match 2.6%; Score 73; DB 9; Length 191652;
Best Local Similarity 100.0%; Pred. No. 1.2e-26;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1983 GCACCTGTAATCCAGCTACTTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGAGG 2042

Db 188583 GCACCTGTAATCCAGCTACTTTGGAGGCTGAGGAGGAGAGTAATCGTTGAACCCAGAGG 188524
Qy 2043 TGGAGGTTGCAGT 2055
Db 188523 TGGAGGTTGCAGT 188511
RESULT 97
AC017100 202950 bp DNA linear PRI 25-JAN-2002
LOCUS Homo sapiens BAC clone RP11-549B18 from 18, complete sequence.
DEFINITION AC017100
ACCESSION AC017100
VERSION AC017100.4 GI:11120958
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 202950)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 202950)
AUTHORS Edwards, J., Gregory, S., Maupin, R. and O'Rourke, K.
TITLE The sequence of Homo sapiens BAC clone RP11-549B18
JOURNAL Unpublished (2002)
REFERENCE 3 (bases 1 to 202950)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Genome Sequencing Center, Washington University, School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 202950)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 202950)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 202950)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Nov 8, 2000 this sequence version replaced gi:7631062.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0549B18

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.


```
QY      2043 TCGAGGTTCCACT 2055
Db      36169 TCGAGGTTCCACT 36181

RESULT 98
AP001336/c
LOCUS   AP001336      216441 bp    DNA    linear    HTG 19-AUG-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-735F7 map 18q12, WORKING
DRAFT SEQUENCE, 18 unordered pieces.
ACCESSION AP001336.3 GI:9857606
VERSION   HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS  Homo sapiens DNA, clone:RP11-735F7.
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   1 (bases 1 to 216441)
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
          Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE     Homo sapiens 216,441 genomic DNA of 18q12
JOURNAL   Published Only in DataBase (2000)
REFERENCE 2 (bases 1 to 216441)
          Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
          Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE     Direct Submission
JOURNAL   Submitted (02-MAR-2000) Masahira Hattori, The Institute of Physical
          and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
          1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
          (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
          Tel.81-42-778-9923, Fax.81-42-778-9924)
          On Aug 19, 2000 this sequence version replaced gl:8117259.
          ----- Genome Center
          Center: RIKEN Genomic Sciences Center(GSC)
          Center code: RIKEN
          Web site: http://hgp.gsc.riken.go.jp/
          Contact: hattori@gsc.riken.go.jp
          ----- Project Information
          Center project name: HumDraft18
          Center clone name: RP11-735F7
          ----- Summary Statistics
          Sequencing vector: PCR products; 100% of reads
          Chemistry: Dye-terminator ET-amersham; 100% of reads
          Assembly program: Phrap; version 0.990329
          Consensus quality: 210841 bases at least Q40
          Consensus quality: 212964 bases at least Q30
          Consensus quality: 214061 bases at least Q20
          Insert size: 214741; sum-of-contigs
          Quality coverage: 10.84x in Q20 bases; sum-of-contigs
          -----
NOTE: This is a 'working draft' sequence. It currently consists of
18 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1      35982 contig of 35982 bp in length
36083  65042 contig of 28960 bp in length
65143  92388 contig of 27246 bp in length
92489  116561 contig of 24073 bp in length
116662 137695 contig of 21034 bp in length
137796 154184 contig of 16389 bp in length
154285 166582 contig of 12298 bp in length
166683 178123 contig of 11441 bp in length
178224 184709 contig of 7022 bp in length
184810 191831 contig of 6486 bp in length
191932 198683 contig of 6752 bp in length
198784 205890 contig of 3730 bp in length
205991 209826 contig of 3277 bp in length
209927 212133 contig of 2207 bp in length
212234 213778 contig of 1545 bp in length

NOTE: This is a 'working draft' sequence. It currently consists of
18 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1      35982 contig of 35982 bp in length
36083  65042 contig of 28960 bp in length
65143  92388 contig of 27246 bp in length
92489  116561 contig of 24073 bp in length
116662 137695 contig of 21034 bp in length
137796 154184 contig of 16389 bp in length
154285 166582 contig of 12298 bp in length
166683 178123 contig of 11441 bp in length
178224 184709 contig of 7022 bp in length
184810 191831 contig of 6486 bp in length
191932 198683 contig of 6752 bp in length
198784 205890 contig of 3730 bp in length
205991 209826 contig of 3277 bp in length
209927 212133 contig of 2207 bp in length
212234 213778 contig of 1545 bp in length
```

```
213879 215264 contig of 1386 bp in length
215365 216441 contig of 1077 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1      35982: contig of 35982 bp in length
35983 36082: gap of 100 bp
36083 65042: contig of 28960 bp in length
65043 65142: gap of 100 bp
65143 92388: contig of 27246 bp in length
92389 92488: gap of 100 bp
92489 116561: contig of 24073 bp in length
116562 116661: gap of 100 bp
116662 137695: contig of 21034 bp in length
137696 137795: gap of 100 bp
137796 154184: contig of 16389 bp in length
154185 154284: gap of 100 bp
154285 166582: contig of 12298 bp in length
166583 166682: gap of 100 bp
166683 178123: contig of 11441 bp in length
178124 178223: gap of 100 bp
178224 184709: contig of 6486 bp in length
184710 184809: gap of 100 bp
184810 191831: contig of 7022 bp in length
191832 191931: gap of 100 bp
191932 198683: contig of 6752 bp in length
198684 198783: gap of 100 bp
198784 202513: contig of 3730 bp in length
202514 202613: gap of 100 bp
202614 205890: contig of 3277 bp in length
205891 205990: gap of 100 bp
205991 209826: contig of 3836 bp in length
209827 209926: gap of 100 bp
209927 212133: contig of 2207 bp in length
212134 212233: gap of 100 bp
212234 213778: contig of 1545 bp in length
213779 213878: gap of 100 bp
213879 215264: contig of 1386 bp in length
215265 215364: gap of 100 bp
215365 216441: contig of 1077 bp in length.
FEATURES
Location/Qualifiers
1..216441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18q12"
/clone="RP11-735F7"
misc_feature
1..35982
/note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature
36083..65042
/note="assembly_fragment"
65143..92388
/note="assembly_fragment"
92489..116561
/note="assembly_fragment"
116662..137695
/note="assembly_fragment"
137796..154184
/note="assembly_fragment"
154285..166582
/note="assembly_fragment"
166683..178123
/note="assembly_fragment"
178224..184709
/note="assembly_fragment clone_end:T7 vector_side:left"
184810..191831
/note="assembly_fragment"
191932..198683
```

```
/note="assembly_fragment"
198784..202513
/note="assembly_fragment"
202614..205890
/note="assembly_fragment"
205991..209826
/note="assembly_fragment"
209927..212133
/note="assembly_fragment"
212234..213778
/note="assembly_fragment"
213879..215264
/note="assembly_fragment"
215365..216441
/note="assembly_fragment"
BASE COUNT 63565 a 42207 c 43579 g 65390 t 1700 others
ORIGIN

Query Match      2.6%; Score 73; DB 2; Length 216441;
Best Local Similarity 100.0%; Pred. No. 1.1e-26;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1983 GCACCTGTATCCAGCTACTTGGAGGCTGAGCAGGAGATCGCTTGAACCCAGGAG 2042
      |||||
Db 124755 GCACCTGTATCCAGCTACTTGGAGGCTGAGCAGGAGATCGCTTGAACCCAGGAG 124696
      |||||

QY 2043 TGGAGGTTGCAGT 2055
      |||||
Db 124695 TGGAGGTTGCAGT 124683

RESULT 99
AC022460/c
LOCUS
DEFINITION Homo sapiens chromosome 12 clone RP11-403N24, WORKING DRAFT
ACCESSION AC022460
VERSION AC022460.4 GI:8954357
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 239008)
Waterston,R.H.
Unpublished
The sequence of Homo sapiens clone
REFERENCE 2 (bases 1 to 239008)
Waterston,R.H.
Direct Submission
Submitted (03-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 7, 2000 this sequence version replaced gi:8570009.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NR0403N24
----- Summary Statistics -----
Sequencing vector: M13; 75%
Chemistry: Dye-terminator; 25%
Assembly: Dye-terminator Big Dye; 25% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 223186 bases at least Q40
Consensus quality: 228280 bases at least Q30
Consensus quality: 230978 bases at least Q20
Insert size: 212000; agarose-gel
Insert size: 236308; sum-of-contigs
Quality coverage: 5.72 in Q20 bases; agarose-gel
```

```
Quality coverage: 5.64 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1012: contig of 1012 bp in length
* 1013 1112: gap of unknown length
* 1113 2339: contig of 1227 bp in length
* 2340 2439: gap of unknown length
* 2440 3628: contig of 1189 bp in length
* 3629 3728: gap of unknown length
* 3729 4968: contig of 1240 bp in length
* 4969 5068: gap of unknown length
* 5069 6838: contig of 1769 bp in length
* 6839 8331: gap of unknown length
* 8332 8432: contig of 1394 bp in length
* 8433 9462: gap of unknown length
* 9463 12825: contig of 1030 bp in length
* 12826 12825: gap of unknown length
* 12826 14654: contig of 3264 bp in length
* 14655 14754: gap of unknown length
* 14755 16752: contig of 1728 bp in length
* 16753 16851: gap of unknown length
* 16852 19355: contig of 1998 bp in length
* 19356 19455: gap of unknown length
* 19456 22127: contig of 2504 bp in length
* 22128 22228: gap of unknown length
* 22229 26873: contig of 2672 bp in length
* 26874 26973: gap of unknown length
* 26974 32655: contig of 4646 bp in length
* 32656 32755: gap of unknown length
* 32756 38338: contig of 5681 bp in length
* 38339 38438: gap of unknown length
* 38439 47366: contig of 5584 bp in length
* 47367 47467: gap of unknown length
* 47468 55621: contig of 8928 bp in length
* 55622 61872: contig of unknown length
* 61873 61972: gap of unknown length
* 61973 68095: contig of 8054 bp in length
* 68096 68195: gap of unknown length
* 68196 75556: contig of 6123 bp in length
* 75557 75556: gap of unknown length
* 75557 83079: contig of 7360 bp in length
* 83080 83179: contig of unknown length
* 83180 95044: gap of unknown length
* 95045 95144: contig of 11865 bp in length
* 95145 107247: gap of unknown length
* 107248 107347: contig of 12103 bp in length
* 107348 121561: contig of unknown length
* 121562 121661: gap of unknown length
* 121662 138129: contig of 14214 bp in length
* 138130 138229: contig of 16468 bp in length
* 138230 156911: gap of unknown length
* 156912 157011: contig of 18682 bp in length
* 157012 184159: gap of unknown length
* 184160 184259: contig of 27148 bp in length
* 184260 239008: gap of unknown length
* 239009 54749 bp in length.
Location/Qualifiers
1. 239008
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-403N24"
```

FEATURES
Source

BASE COUNT 59374 a 57881 c 58172 g 60872 t 2709 others
ORIGIN

Query Match 2.6%; Score 73; DB 2; Length 239008;
Best Local Similarity 100.0%; Pred. No. 1.1e-26;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1983 GCACCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGTAATCGCTTGAACCCAGGAGG 2042
|||||
DB 61157 GCACCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGTAATCGCTTGAACCCAGGAGG 61098
|||||

OY 2043 TGGAGGTTGCAGT 2055
|||||
DB 61097 TGGAGGTTGCAGT 61085
|||||

Search completed: July 18, 2003, 01:11:12
Job time : 7146 secs

Db 42497 CACCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGTAATCGCTTGAACCCAGGAGGT 42556
OY 2044 GGAGGTTGCAGT 2055
|||||
DB 42557 GGAGGTTGCAGT 42568
|||||

RESULT 100
AL669821
LOCUS 89211 bp DNA linear PRI 24-APR-2002
DEFINITION Human DNA sequence from clone Xbac-116A1 on chromosome 6, complete
sequence.
ACCESSION AL669821
VERSION AL669821.9 GI:20152403
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1
Mashreghi-Mohammadi, M.
Direct Submission

Submitted (23-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Apr 15, 2002 this sequence version replaced gi:20145365.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30);
an attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep Xbac-116A1 is
from a CHORI-501 human bac - PGF cell line library VECTOR:

PTARBAC2.1
This sequence was generated from part of bacterial clone contigs
constructed by the MHC Haplotype Consortium and collaborators.
Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6/MHC.

FEATURES
Location/Qualifiers
1..89211
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="XXbac-116A1"
/clone_lib="CHORI-501"
BASE COUNT 26990 a 19373 c 19477 g 23371 t
ORIGIN

Query Match 2.6%; Score 72; DB 9; Length 89211;
Best Local Similarity 100.0%; Pred. No. 4.1e-26;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1984 CACCTGTAATCCAGCTACTTGGAGGCTGAGGAGGAGTAATCGCTTGAACCCAGGAGGT 2043
|||||

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 17, 2003, 20:58:25 ; Search time 1772 seconds
(without alignments)
3251.894 Million cell updates/sec

Title: US-09-966-880A-8

Perfect score: 198

Sequence: 1 MDSLLMNRKRKFLYQPKNVRW.....ILLPLYEVDLDRFAFTGL 198

Scoring table:
OLIGO Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4106490

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Command line parameters: -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09966880/runat_16072003_115617_17474/app_query.fasta_1.391
-DB=GenEmbl -QFWT=fastap -SUFFIX=10align.rge -MINMATCH=0.1 -LOOPCT=0
-LOOPEXT=0 -DUALS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=500 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=100 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880@cgn_1.1 3745 @runat_16072003_115617_17474 -NCPU=6 -ICPU=3
-NO_MWAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl :
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_ov.*
5: gb_om.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pt.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_hgt_hum.*
31: em_hgt_inv.*
32: em_hgt_other.*
33: em_hgt_mus.*
34: em_hgt_pln.*
35: em_hgt_rod.*
36: em_hgt_mam.*
37: em_hgt_vrt.*
38: em_sy.*
39: em_hgtgo_hum.*
40: em_hgtgo_mus.*
41: em_hgtgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	198	100.0	596	9 AF529823	AF529823 Homo sapi
2	198	100.0	596	9 AF529826	AF529826 Homo sapi
3	198	100.0	596	12 AF529837	AF529837 Mus muscu
4	198	100.0	1837	9 BC006296	BC006296 Homo sapi
5	198	100.0	2791	9 AB040431	AB040431 Homo sapi
6	195	98.5	596	12 AF529829	AF529829 Mus muscu
7	195	98.5	597	12 AF529842	AF529842 Crictetulu
8	195	98.5	597	12 AF529843	AF529843 Crictetulu
9	195	98.5	597	12 AF529844	AF529844 Crictetulu
10	195	98.5	597	12 AF529847	AF529847 Crictetulu
11	195	98.5	597	12 AF529852	AF529852 Crictetulu
12	190	96.0	597	12 AF529846	AF529846 Crictetulu
13	177	89.4	596	12 AF529832	AF529832 Mus muscu
14	175	88.4	596	9 AF529815	AF529815 Homo sapi
15	175	88.4	596	12 AF529831	AF529831 Mus muscu
16	172	86.9	597	12 AF529851	AF529851 Crictetulu
17	171	86.4	596	9 AF529825	AF529825 Homo sapi
18	170	85.9	591	12 AF529855	AF529855 Crictetulu
19	169	85.4	596	12 AF529833	AF529833 Mus muscu
20	169	85.4	596	12 AF529835	AF529835 Mus muscu
21	162	81.8	547	12 AF529856	AF529856 Crictetulu
22	160	80.8	596	9 AF529822	AF529822 Homo sapi
23	149	75.3	595	12 AF529834	AF529834 Mus muscu
24	146	73.7	596	12 AF529836	AF529836 Mus muscu
25	146	73.7	597	12 AF529853	AF529853 Crictetulu
26	144	72.7	597	12 AF529828	AF529828 Mus muscu
27	143	72.2	596	9 AF529824	AF529824 Homo sapi
28	140	70.7	597	12 AF529841	AF529841 Crictetulu
29	139	70.2	596	12 AF529830	AF529830 Mus muscu
30	136	68.7	596	12 AF529838	AF529838 Mus muscu
31	135	68.2	597	12 AF529845	AF529845 Crictetulu
32	128	64.6	596	9 AF529821	AF529821 Homo sapi
33	126	63.6	597	12 AF529848	AF529848 Crictetulu
34	122	61.6	596	12 AF529839	AF529839 Mus muscu
35	122	61.6	597	12 AF529850	AF529850 Crictetulu
36	118	59.6	596	9 AF529818	AF529818 Homo sapi
37	118	59.6	596	9 AF529819	AF529819 Homo sapi
38	116	58.6	596	9 AF529820	AF529820 Homo sapi
39	116	58.6	597	12 AF529849	AF529849 Crictetulu
40	115	58.1	577	12 AF529840	AF529840 Mus muscu
41	114	57.6	596	9 AF529816	AF529816 Homo sapi
42	109	55.1	596	9 AF529827	AF529827 Homo sapi
43	108	54.5	597	12 AF529854	AF529854 Crictetulu
44	96	48.5	596	9 AF529817	AF529817 Homo sapi
45	90	45.5	11204	9 AB040430	AB040430 Homo sapi
46	90	45.5	71132	9 AC092184	AC092184 Homo sapi
47	42	21.2	2440	10 AF132979	AF132979 Mus muscu
48	42	21.2	143044	2 AC094826	AC094826 Rattus no
49	42	21.2	177103	2 AC120617	AC120617 Rattus no
50	38	19.2	50269	2 AC119975	AC119975 Mus muscu

C 51	34	17.2	98161	2	AC109119	AC109119 Rattus no	124	8	4.0	2086	8	AF224075	AF224075 Solanum n
C 52	16	8.1	50269	2	AC119975	Mus muscu	125	8	4.0	2086	8	CBU08916	CBU08916 Capsicum ba
C 53	10	5.1	67244	2	AC068309	Mus muscu	126	8	4.0	2086	8	CBU47428	CBU47428 Cyphomandra
C 54	9	4.5	7750	3	DMSPALTR	Y07653 D.melanog	127	8	4.0	2086	8	DSU08917	DSU08917 Datura stra
C 55	9	4.5	10945	1	AE007918	AE007918 Agrobacte	128	8	4.0	2086	8	JAU08919	JAU08919 Juanelloa a
C 56	9	4.5	11013	1	AE008970	AE008970 Agrobacte	129	8	4.0	2086	8	JPU47429	JPU47429 Jalomata p
C 57	9	4.5	59915	2	AC100984	Mus muscu	130	8	4.0	2086	8	LCU08920	LCU08920 Lycium cest
C 58	9	4.5	103746	2	AC014357	AC014357 Drosophil	131	8	4.0	2086	8	LEU08921	LEU08921 Lycium cest
C 59	9	4.5	141899	2	AC087457	AC087457 Homo sapi	132	8	4.0	2086	8	LEU73797	LEU73797 Lycianthes
C 60	9	4.5	150147	2	AC123206	AC123206 Rattus no	133	8	4.0	2086	8	MOU08922	MOU08922 Mandruora
C 61	9	4.5	154288	2	AL513164	AL513164 Human DNA	134	8	4.0	2086	8	NAU08923	NAU08923 Nicotiana a
C 62	9	4.5	158016	2	AC113662	AC113662 Rattus no	135	8	4.0	2086	8	NSU08924	NSU08924 Nicotiana ph
C 63	9	4.5	173509	2	AP002436	AP002436 Homo sapi	136	8	4.0	2086	8	PAU08925	PAU08925 Nolana spat
C 64	9	4.5	176053	2	AC068190	AC068190 Homo sapi	137	8	4.0	2086	8	POTCP1A	POTCP1A Physalis al
C 65	9	4.5	179510	2	AC127421	AC127421 Mus muscu	138	8	4.0	2086	8	SAU47417	SAU47417 Solanum arb
C 66	9	4.5	180668	2	AC020857	AC020857 Mus muscu	139	8	4.0	2086	8	SAU47423	SAU47423 Solanum arb
C 67	9	4.5	187948	2	AP003402	AP003402 Homo sapi	140	8	4.0	2086	8	SAU72752	SAU72752 Solanum arg
C 68	9	4.5	188067	3	AC007083	AC007083 Drosophil	141	8	4.0	2086	8	SCU072751	SCU072751 Solanum cor
C 69	9	4.5	188136	2	AL840638	AL840638 Dario rer	142	8	4.0	2086	8	SGU08930	SGU08930 Solandra gr
C 70	9	4.5	190960	2	AC036188	AC036188 Homo sapi	143	8	4.0	2086	8	SLU72749	SLU72749 Solanum gla
C 71	9	4.5	199386	8	NCB14D6	NCB14D6	144	8	4.0	2086	8	SPU47421	SPU47421 Solanum lut
C 72	9	4.5	207584	2	AC103892	AC103892 Rattus no	145	8	4.0	2086	8	SPU47422	SPU47422 Solanum phy
C 73	9	4.5	232951	2	AC091322	AC091322 Mus muscu	146	8	4.0	2086	8	SRU47424	SRU47424 Solanum pse
C 74	9	4.5	244237	3	AE003632	AE003632 Drosophil	147	8	4.0	2086	8	SSU47425	SSU47425 Solanum tos
C 75	8	4.0	245	8	TCARITS1	TCARITS1	148	8	4.0	2086	8	STU72750	STU72750 Solanum sea
C 76	8	4.0	354	9	HS2479528	HS2479528 Homo sapi	149	8	4.0	2086	8	SWU47426	SWU47426 Solanum tri
C 77	8	4.0	500	6	AX380702	AX380702 Sequence	150	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 78	8	4.0	569	11	HC070H12	HC070H12	151	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 79	8	4.0	653	11	PM3C3B	PM3C3B	152	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 80	8	4.0	984	6	A48781	A48781 Sequence 8	153	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 81	8	4.0	984	6	A58739	A58739 Sequence 2	154	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 82	8	4.0	984	6	A8018707	A8018707 Sequence	155	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 83	8	4.0	1006	6	A48778	A48778 Sequence 5	156	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 84	8	4.0	1006	6	A8018705	A8018705 Sequence	157	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 85	8	4.0	1101	3	AY094934	AY094934 Drosophil	158	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 86	8	4.0	1149	6	A48779	A48779 Sequence 6	159	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 87	8	4.0	1149	6	A48782	A48782 Sequence 9	160	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 88	8	4.0	1149	6	AR018706	AR018706 Sequence	161	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 89	8	4.0	1149	6	AR018708	AR018708 Sequence	162	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 90	8	4.0	1188	6	A58738	A58738 Sequence 1	163	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 91	8	4.0	1230	6	A58740	A58740 Sequence 3	164	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 92	8	4.0	1237	8	AF238055	AF238055 Hydrastis	165	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 93	8	4.0	1465	1	AF060158	AF060158 Aucuba ja	166	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 94	8	4.0	1465	1	AF071567	AF071567 Bordetell	167	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 95	8	4.0	1565	8	EANDHF1	EANDHF1	168	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 96	8	4.0	1857	1	AF353205	AF353205 Exocarp	169	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 97	8	4.0	1886	8	AK094292	AK094292 Azospiril	170	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 98	8	4.0	1938	8	HAU12660	HAU12660 Hemigraphis	171	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 99	8	4.0	1938	8	AF257485	AF257485 Goyazia r	172	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 100	8	4.0	2025	8	ESU12659	ESU12659 Eremomastax	173	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 101	8	4.0	2052	8	AF206197	AF206197 Nematanth	174	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 102	8	4.0	2057	8	AF013689	AF013689 Nematanth	175	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 103	8	4.0	2057	8	AF206198	AF206198 Nematanth	176	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 104	8	4.0	2063	8	AF206196	AF206196 Codonanth	177	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 105	8	4.0	2067	8	AF176635	AF176635 Tylopsaca	178	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 106	8	4.0	2071	8	AF224068	AF224068 Solanum m	179	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 107	8	4.0	2077	8	AF040144	AF040144 Bellonia	180	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 108	8	4.0	2080	8	SAU47415	SAU47415 Solanum abu	181	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 109	8	4.0	2084	8	SLU47420	SLU47420 Solanum lac	182	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 110	8	4.0	2086	8	ABU08915	ABU08915 Atropa bell	183	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 111	8	4.0	2086	8	AF224061	AF224061 Solanum a	184	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 112	8	4.0	2086	8	AF224062	AF224062 Solanum a	185	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 113	8	4.0	2086	8	AF224063	AF224063 Normania	186	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 114	8	4.0	2086	8	AF224064	AF224064 Solanum p	187	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 115	8	4.0	2086	8	AF224065	AF224065 Triguera	188	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 116	8	4.0	2086	8	AF224066	AF224066 Solanum v	189	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 117	8	4.0	2086	8	AF224067	AF224067 Solanum e	190	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 118	8	4.0	2086	8	AF224069	AF224069 Solanum m	191	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 119	8	4.0	2086	8	AF224070	AF224070 Solanum v	192	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 120	8	4.0	2086	8	AF224071	AF224071 Solanum c	193	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 121	8	4.0	2086	8	AF224072	AF224072 Solanum c	194	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 122	8	4.0	2086	8	AF224073	AF224073 Solanum j	195	8	4.0	2086	8	WU72755	WU72755 Solanum wal
C 123	8	4.0	2086	8	AF224074	AF224074 Solanum m	196	8	4.0	2086	8	WU72755	WU72755 Solanum wal

c 197	8	4.0	13483	8	AF098806	Sorghum b	270	8	4.0	112799	2	AC117860	AC117860	Rattus no
c 198	8	4.0	13782	9	AC108042	Homo sapi	271	8	4.0	112910	2	AC131019	AC131019	Rattus no
c 199	8	4.0	13994	1	AE005952	Caulobact	272	8	4.0	113333	2	AC096958	AC096958	Rattus no
c 200	8	4.0	18120	9	HSB20F6	Human DNA s	c 273	8	4.0	113759	9	AC008844	AC008844	Homo sapi
c 201	8	4.0	20328	2	AC014603	Drosophill	274	8	4.0	114870	2	AC010755	AC010755	Homo sapi
c 202	8	4.0	25951	5	AB004829	Fugu rubr	275	8	4.0	115199	9	AC003695	AC003695	Homo sapi
c 203	8	4.0	26590	9	AC090672	Homo sapi	c 276	8	4.0	115905	9	HS57E3	AC094266	Rattus no
c 204	8	4.0	28170	6	AX283721	Sequence	c 277	8	4.0	116164	2	AC094338	AC094338	Rattus no
c 205	8	4.0	31664	3	CEF10A3	Z92829 Caenorhabdi	c 278	8	4.0	116356	2	OSJN00125	OSJN00125	Rattus no
c 206	8	4.0	32798	6	AX382187	Sequence	c 279	8	4.0	120558	8	AP005479	AP005479	Oryza sat
c 207	8	4.0	33007	12	AF323988	EGFP expr	c 280	8	4.0	120710	8	AC002396	AC002396	Arabidops
c 208	8	4.0	33592	6	AX084504	Sequence	c 281	8	4.0	122358	8	AC002396	AC002396	Arabidops
c 209	8	4.0	33699	6	AX084506	Sequence	c 282	8	4.0	122466	9	HSJ841K13	HSJ841K13	Human DNA
c 210	8	4.0	33988	6	AX084517	Sequence	c 283	8	4.0	122580	2	AC128369	AC128369	Rattus no
c 211	8	4.0	34303	6	AR091536	Sequence	c 284	8	4.0	125866	8	OSJN00283	OSJN00283	Rattus no
c 212	8	4.0	34303	6	AR102229	Sequence	c 285	8	4.0	126734	8	AP003849	AP003849	Oryza sat
c 213	8	4.0	34341	6	AX084505	Sequence	c 286	8	4.0	127829	2	AP004154	AP004154	Oryza sat
c 214	8	4.0	34448	6	AX084507	Sequence	c 287	8	4.0	129311	8	AP003218	AP003218	Oryza sat
c 215	8	4.0	34737	6	AX084518	Sequence	c 288	8	4.0	129420	8	AC087547	AC087547	Oryza sat
c 216	8	4.0	35408	6	AR163568	Sequence	c 289	8	4.0	132581	2	AP003742	AP003742	Oryza sat
c 217	8	4.0	35408	6	AR166442	Sequence	c 290	8	4.0	135469	2	AL136127	AL136127	Homo sapi
c 218	8	4.0	35724	6	AX084516	Sequence	c 291	8	4.0	136116	2	AC110874	AC110874	Gallus ga
c 219	8	4.0	35764	12	AY046510	AY046510 Adenovira	c 292	8	4.0	136917	2	AP005097	AP005097	Oryza sat
c 220	8	4.0	35935	6	AR091533	Sequence	c 293	8	4.0	138589	9	AC092966	AC092966	Homo sapi
c 221	8	4.0	35935	6	AR102226	Sequence	c 294	8	4.0	138724	2	AC114239	AC114239	Rattus no
c 222	8	4.0	35935	6	AR116313	Sequence	c 295	8	4.0	139147	2	AC006725	AC006725	Caenorhab
c 223	8	4.0	35935	6	AX451988	Sequence	c 296	8	4.0	139381	2	OSJN01002	OSJN01002	Oryza sat
c 224	8	4.0	35935	14	ADRCMPGEN	W73260 Mastadenovi	c 297	8	4.0	139480	9	HUAC002331	HUAC002331	Homo sapi
c 225	8	4.0	35937	14	ADRCG	JO1917 Adenovirus	c 298	8	4.0	139646	2	AC115934	AC115934	Mus muscu
c 226	8	4.0	36114	6	AX084519	Sequence	c 299	8	4.0	140934	2	AC112553	AC112553	Rattus no
c 227	8	4.0	36154	6	AX468857	Sequence	c 300	8	4.0	141633	8	OSJN00278	OSJN00278	Oryza sat
c 228	8	4.0	36154	6	AX468865	Sequence	c 301	8	4.0	142689	9	HSB348B13	HSB348B13	Human DNA
c 229	8	4.0	37142	2	AC025045	Sequence	c 302	8	4.0	142707	2	AP004338	AP004338	Oryza sat
c 230	8	4.0	37391	6	AX468871	Sequence	c 303	8	4.0	144268	2	OSJN00212	OSJN00212	Oryza sat
c 231	8	4.0	38041	6	AX468869	Sequence	c 304	8	4.0	144514	2	AC115534	AC115534	Sus scrof
c 232	8	4.0	38209	1	SC5G9	AL117385 Streptomy	c 305	8	4.0	144547	9	AC015819	AC015819	Homo sapi
c 233	8	4.0	40897	8	SPCC794	AL023595 S.pombe c	c 306	8	4.0	145467	2	AP004270	AP004270	Oryza sat
c 234	8	4.0	42031	3	CEB35A5	Z71185 Caenorhabdi	c 307	8	4.0	145668	9	AC092272	AC092272	Homo sapi
c 235	8	4.0	42071	1	SC2G61	AL359949 Streptomy	c 308	8	4.0	145859	9	AC009510	AC009510	Homo sapi
c 236	8	4.0	46676	9	AL356490	AL356490 Human DNA	c 309	8	4.0	145887	2	AC091435	AC091435	Homo sapi
c 237	8	4.0	47400	2	AC129582	AC129582 Mus muscu	c 310	8	4.0	146810	2	AL627262	AL627262	Danio rer
c 238	8	4.0	52111	2	AL358796_3	Continuation (4 of	c 311	8	4.0	146921	8	AP002836	AP002836	Oryza sat
c 239	8	4.0	58456	2	AC127514	AC127514 Homo sapi	c 312	8	4.0	147021	9	AC006501	AC006501	Homo sapi
c 240	8	4.0	59000	9	AF314199S1	AF314199 Homo sapi	c 313	8	4.0	147405	2	AC118576	AC118576	Didelphis
c 241	8	4.0	63285	2	AC122785	AC122785 Mus muscu	c 314	8	4.0	150191	8	OSJN00038	OSJN00038	Oryza sat
c 242	8	4.0	63286	8	AC007264	AC007264 Arabidops	c 315	8	4.0	150626	9	HSJ177P10	HSJ177P10	Human DNA
c 243	8	4.0	71506	9	AC092452	AC092452 Homo sapi	c 316	8	4.0	151559	2	AC015863	AC015863	Homo sapi
c 244	8	4.0	73094	2	AL671114	AL671114 Homo sapi	c 317	8	4.0	151819	2	AC108754	AC108754	Oryza sat
c 245	8	4.0	74774	2	AC100316	AC100316 Mus muscu	c 318	8	4.0	152523	9	AC007009	AC007009	Homo sapi
c 246	8	4.0	75892	2	AC101297	AL035470 Mus muscu	c 319	8	4.0	152699	2	AP005125	AP005125	Oryza sat
c 247	8	4.0	80120	9	HS480J14	X90568 H.sapiens m	c 320	8	4.0	153336	2	AC025020	AC025020	Homo sapi
c 248	8	4.0	81940	9	HS7ITIN2B	AC021944 Homo sapi	c 321	8	4.0	153805	9	AC090987	AC090987	Homo sapi
c 249	8	4.0	87551	2	AC021944	AC021944 Homo sapi	c 322	8	4.0	153866	9	AC009307	AC009307	Homo sapi
c 250	8	4.0	88535	2	AC099423	AC099423 Rattus no	c 323	8	4.0	153926	9	AL160059	AL160059	Human DNA
c 251	8	4.0	89370	3	AC084153	AC084153 Caenorhab	c 324	8	4.0	154123	2	AC019562	AC019562	Drosophill
c 252	8	4.0	89473	8	TIK7	AC013427 Sequence	c 325	8	4.0	154153	9	AL445933	AL445933	Human DNA
c 253	8	4.0	91071	8	AC002339	AC002339 Arabidops	c 326	8	4.0	154210	2	AC078888	AC078888	Homo sapi
c 254	8	4.0	93427	9	AL591076	AL591076 Human DNA	c 327	8	4.0	154440	2	AC108176	AC108176	Bos tauru
c 255	8	4.0	93475	5	AL672088	AL672088 Zebrafish	c 328	8	4.0	154625	2	AP005383	AP005383	Oryza sat
c 256	8	4.0	94024	9	AC007368	AC007368 Homo sapi	c 329	8	4.0	154683	9	AC006157	AC006157	Homo sapi
c 257	8	4.0	94682	2	AC128917	AC128917 Rattus no	c 330	8	4.0	154814	9	AC114802	AC114802	Homo sapi
c 258	8	4.0	98980	2	AP004215	AP004215 Oryza sat	c 331	8	4.0	155939	8	CHNTXX	CHNTXX	Nicotiana t
c 259	8	4.0	100000	9	AB020866	AB020866 Homo sapi	c 332	8	4.0	156000	2	AC060815	AC060815	Homo sapi
c 260	8	4.0	100499	9	AL138769	AL138769 Human DNA	c 333	8	4.0	156277	9	AC107015	AC107015	Homo sapi
c 261	8	4.0	105928	2	AC111925	AC111925 Rattus no	c 334	8	4.0	156687	8	ABE316582	ABE316582	Atropa be
c 262	8	4.0	107139	2	AL360223	AL360223 Homo sapi	c 335	8	4.0	156856	2	AC097235	AC097235	Rattus no
c 263	8	4.0	107655	2	AC084840	AC084840 Homo sapi	c 336	8	4.0	156929	2	AC026854	AC026854	Homo sapi
c 264	8	4.0	108799	9	AC090989	AC090989 Homo sapi	c 337	8	4.0	157262	9	AC104827	AC104827	Homo sapi
c 265	8	4.0	110000	2	LMFLCHR34_10	Continuation (2 of	c 338	8	4.0	157926	2	AC096047	AC096047	Rattus no
c 266	8	4.0	110000	2	LMFLCHR36_17	Continuation (11 o	c 339	8	4.0	158078	9	AC005094	AC005094	Homo sapi
c 267	8	4.0	110000	2	LMFLCHR36_17	Continuation (18 o	c 340	8	4.0	158133	8	AP002816	AP002816	Oryza sat
c 268	8	4.0	110000	2	AC026388_3	Continuation (4 of	c 341	8	4.0	158242	3	AC115483	AC115483	Drosophill
c 269	8	4.0	110042	9	HS179E3	Z95113 Human DNA s	c 342	8	4.0	158349	9	AC024908	AC024908	Homo sapi

c 343	8	4.0	158719	2	AC112081	AC112081	Rattus no	c 416	8	4.0	183990	8	AF114171	AF114171	Sorghum b
344	8	4.0	159132	2	AC118378	AC18378	Rattus no	c 417	8	4.0	184203	10	AL669913	AL669913	Mouse DNA
345	8	4.0	159694	2	AC098586	AC098586	Homo sapi	c 418	8	4.0	184412	2	AC109147	AC109147	Mus muscu
346	8	4.0	159738	9	AC073352	AC073352	Homo sapi	c 419	8	4.0	185050	2	AC127144	AC127144	Rattus no
347	8	4.0	160557	9	AC117945	AC117945	Homo sapi	c 420	8	4.0	185334	2	AC129392	AC129392	Rattus no
c 348	8	4.0	161247	9	AC069304	AC069304	Homo sapi	c 421	8	4.0	185635	2	AC122372	AC122372	Mus muscu
349	8	4.0	161437	2	AC093434	AC093434	Homo sapi	c 422	8	4.0	185741	9	AC006203	AC006203	Homo sapi
c 350	8	4.0	161987	9	AC006070	AC006070	Homo sapi	c 423	8	4.0	186760	2	AC026056	AC026056	Homo sapi
351	8	4.0	162504	2	AC106209	AC106209	Rattus no	c 424	8	4.0	187046	2	AC079912	AC079912	Mus muscu
c 352	8	4.0	162573	2	AC115535	AC115535	Sus scrof	c 425	8	4.0	187508	9	AC104078	AC104078	Homo sapi
353	8	4.0	162712	2	AC026096	AC026096	Homo sapi	c 426	8	4.0	187778	2	AC034166	AC034166	Homo sapi
c 354	8	4.0	162860	2	AC119097	AC119097	Rattus no	c 427	8	4.0	187936	9	AC080003	AC080003	Homo sapi
c 355	8	4.0	163282	2	AC016438	AC016438	Homo sapi	c 428	8	4.0	188324	2	AC087625	AC087625	Homo sapi
356	8	4.0	163410	9	AC006198	AC006198	Homo sapi	c 429	8	4.0	188359	9	AC006369	AC006369	Homo sapi
c 357	8	4.0	163932	2	AC114160	AC114160	Rattus no	c 430	8	4.0	188493	2	AC124181	AC124181	Mus muscu
c 358	8	4.0	164259	2	AC113514	AC113514	Mus muscu	c 431	8	4.0	188507	2	AL844159	AL844159	Mus muscu
c 359	8	4.0	164331	9	AC016254	AC016254	Homo sapi	c 432	8	4.0	189050	1	AL646066	AL646066	Rattus no
c 360	8	4.0	164958	2	AC092729	AC092729	Canis fam	c 433	8	4.0	189576	2	AL772183	AL772183	Mus muscu
c 361	8	4.0	165228	9	AC005325	AC005325	Homo sapi	c 434	8	4.0	189838	2	AC090121	AC090121	Mus muscu
c 362	8	4.0	165497	8	AP003232	AP003232	Oryza sat	c 435	8	4.0	190050	1	AL646059	AL646059	Rattus no
c 363	8	4.0	165866	2	AC087273	AC087273	Homo sapi	c 436	8	4.0	190212	2	AC127097	AC127097	Rattus no
c 364	8	4.0	166474	2	AC107431	AC107431	Rattus no	c 437	8	4.0	190241	2	AC119764	AC119764	Rattus no
c 365	8	4.0	166517	9	AL356672	AL356672	Human DNA	c 438	8	4.0	190322	9	AC093124	AC093124	Papio cyn
c 366	8	4.0	168862	9	AC079880	AC079880	Homo sapi	c 439	8	4.0	190417	2	AC090319	AC090319	Homo sapi
c 367	8	4.0	167503	2	AC129344	AC129344	Rattus no	c 440	8	4.0	190462	9	AC092535	AC092535	Homo sapi
c 368	8	4.0	167627	9	AC079382	AC079382	Homo sapi	c 441	8	4.0	190624	2	AC013811	AC013811	Homo sapi
c 369	8	4.0	168367	9	AC104440	AC104440	Homo sapi	c 442	8	4.0	190651	9	AP000781	AP000781	Homo sapi
c 370	8	4.0	168407	2	AC090786	AC090786	Homo sapi	c 443	8	4.0	191098	2	AC073518	AC073518	Homo sapi
c 371	8	4.0	169462	2	AC024033	AC024033	Homo sapi	c 444	8	4.0	191170	9	AC006112	AC006112	Homo sapi
c 372	8	4.0	169747	9	AL355352	AL355352	Human DNA	c 445	8	4.0	191667	2	AC124202	AC124202	Mus muscu
c 373	8	4.0	170711	2	AC032020	AC032020	Homo sapi	c 446	8	4.0	192126	9	DJ270M14	DJ270M14	Mus muscu
c 374	8	4.0	170761	9	AC021566	AC021566	Homo sapi	c 447	8	4.0	192130	2	AC116768	AC116768	Mus muscu
c 375	8	4.0	171339	9	AC107904	AC107904	Homo sapi	c 448	8	4.0	192441	2	AC073404	AC073404	Homo sapi
c 376	8	4.0	171676	9	AC120613	AC120613	Rattus no	c 449	8	4.0	192809	2	AC110518	AC110518	Mus muscu
c 377	8	4.0	173056	9	AC006370	AC006370	Homo sapi	c 450	8	4.0	193495	2	AC108780	AC108780	Mus muscu
c 378	8	4.0	173146	2	AC113058	AC113058	Homo sapi	c 451	8	4.0	194206	9	AC021863	AC021863	Homo sapi
c 379	8	4.0	173185	2	AC016399	AC016399	Homo sapi	c 452	8	4.0	196606	9	AP000779	AP000779	Homo sapi
c 380	8	4.0	173239	9	AC117925	AC117925	Rattus no	c 453	8	4.0	197176	2	AC022763	AC022763	Homo sapi
381	8	4.0	173422	2	AC008756	AC008756	Homo sapi	c 454	8	4.0	198326	2	AC087839	AC087839	Oryza sat
c 382	8	4.0	173522	9	AC025904	AC025904	Homo sapi	c 455	8	4.0	198729	2	AC113292	AC113292	Mus muscu
c 383	8	4.0	174032	2	AC073956	AC073956	Homo sapi	c 456	8	4.0	198748	2	AC016719	AC016719	Homo sapi
c 384	8	4.0	174204	2	AC016105	AC016105	Homo sapi	c 457	8	4.0	198929	2	AC101848	AC101848	Mus muscu
c 385	8	4.0	174228	2	AC073638	AC073638	Homo sapi	c 458	8	4.0	199101	10	AL672034	AL672034	Mouse DNA
c 386	8	4.0	174694	9	AC007217	AC007217	Homo sapi	c 459	8	4.0	199287	2	AC102672	AC102672	Mus muscu
c 387	8	4.0	175111	2	AC101999	AC101999	Mus muscu	c 460	8	4.0	199393	2	AC013374	AC013374	Homo sapi
c 388	8	4.0	176096	2	AC111117	AC111117	Mus muscu	c 461	8	4.0	199534	2	AC025959	AC025959	Homo sapi
c 389	8	4.0	176177	2	AC115409	AC115409	Rattus no	c 462	8	4.0	200131	9	AC105213	AC105213	Homo sapi
c 390	8	4.0	176226	2	AC073638	AC073638	Homo sapi	c 463	8	4.0	200240	2	AC011618	AC011618	Homo sapi
c 391	8	4.0	176459	2	AC108684	AC108684	Homo sapi	c 464	8	4.0	200336	2	AC112955	AC112955	Mus muscu
c 392	8	4.0	176525	2	AL845474	AL845474	Mus muscu	c 465	8	4.0	201345	2	AL645912	AL645912	Mus muscu
c 393	8	4.0	176624	2	AC104394	AC104394	Homo sapi	c 466	8	4.0	201780	2	AC124354	AC124354	Mus muscu
c 394	8	4.0	176841	2	AC117096	AC117096	Rattus no	c 467	8	4.0	201802	2	AC124473	AC124473	Mus muscu
c 395	8	4.0	177145	2	AC113156	AC113156	Homo sapi	c 468	8	4.0	201982	2	AC027297	AC027297	Mus muscu
c 396	8	4.0	177476	9	AC010680	AC010680	Homo sapi	c 469	8	4.0	202134	2	AL772249	AL772249	Mus muscu
c 397	8	4.0	177954	9	AC011366	AC011366	Homo sapi	c 470	8	4.0	202163	2	AF235099	AF235099	Homo sapi
c 398	8	4.0	178019	3	AC091206	AC091206	Drosophil	c 471	8	4.0	202696	2	AP001654	AP001654	Homo sapi
c 399	8	4.0	178825	2	AP001203	AP001203	Homo sapi	c 472	8	4.0	204696	2	AC025963	AC025963	Homo sapi
c 400	8	4.0	179094	2	AP004138	AP004138	Oryza sat	c 473	8	4.0	205195	9	AC008546	AC008546	Homo sapi
c 401	8	4.0	179161	2	AC122412	AC122412	Mus muscu	c 474	8	4.0	205954	9	AC008546	AC008546	Homo sapi
c 402	8	4.0	179335	9	AP000870	AP000870	Homo sapi	c 475	8	4.0	206909	9	AC016397	AC016397	Homo sapi
c 403	8	4.0	180000	2	AC004578	AC004578	Homo sapi	c 476	8	4.0	209630	2	AC118019	AC118019	Mus muscu
c 404	8	4.0	180062	2	AC118333	AC118333	Rattus no	c 477	8	4.0	209923	2	AC121806	AC121806	Mus muscu
c 405	8	4.0	180155	2	AL161735	AL161735	Homo sapi	c 478	8	4.0	210296	2	AC121971	AC121971	Mus muscu
c 406	8	4.0	180484	9	AC026759	AC026759	Oryza sat	c 479	8	4.0	212080	2	AC091457	AC091457	Mus muscu
c 407	8	4.0	180619	9	AL356513	AL356513	Human DNA	c 480	8	4.0	213800	2	AC127415	AC127415	Mus muscu
c 408	8	4.0	180829	2	AC022508	AC022508	Homo sapi	c 481	8	4.0	217501	2	AC116128	AC116128	Mus muscu
c 409	8	4.0	180859	2	AC091126	AC091126	Drosophil	c 482	8	4.0	218648	2	AL731708	AL731708	Mus muscu
c 410	8	4.0	180968	2	AC078996	AC078996	Mus muscu	c 483	8	4.0	219038	2	AC083890	AC083890	Mus muscu
c 411	8	4.0	181415	2	AP005195	AP005195	Oryza sat	c 484	8	4.0	219447	9	HUMFLNGGPD	HUMFLNGGPD	Homo sapi
c 412	8	4.0	181632	2	AC053508	AC053508	Homo sapi	c 485	8	4.0	221978	2	AC103664	AC103664	Mus muscu
c 413	8	4.0	181805	10	AL670660	AL670660	Mouse DNA	c 486	8	4.0	222606	2	AC125407	AC125407	Mus muscu
c 414	8	4.0	182440	9	AC096755	AC096755	Homo sapi	c 487	8	4.0	222618	2	AC103335	AC103335	Rattus no
c 415	8	4.0	183118	2	AL355528	AL355528	Homo sapi	c 488	8	4.0	223717	2	AC108896	AC108896	Bos tauru
													AL845264	AL845264	Mus muscu


```

489      8 4.0 228659 2 AC095481 AC095481 Rattus no
490      8 4.0 228856 2 AC087420 AC087420 Mus muscu
491      8 4.0 230851 2 AC115440 AC115440 Rattus no
c 492      8 4.0 231562 9 AC010148 AC010148 Homo sapi
c 493      8 4.0 231772 2 AC125217 AC125217 Mus muscu
c 494      8 4.0 234096 2 AC121818 AC121818 Mus muscu
495      8 4.0 235086 2 AC100043 AC100043 Mus muscu
496      8 4.0 239925 2 AC110921 AC110921 Homo sapi
c 497      8 4.0 247629 2 AC123124 AC123124 Rattus no
c 498      8 4.0 248160 2 AC122000 AC122000 Mus muscu
c 499      8 4.0 257000 1 AP000002 AP000002 Pyrococcu
c 500      8 4.0 257817 2 AC006909 AC006909 Caenorhab

```

ALIGNMENTS

```

RESULT 1
AF529823
LOCUS      Homo sapiens clone Ramos 9 AID (AID) mRNA, partial cds.
DEFINITION
ACCESSION AF529823
VERSION    AF529823.1 GI:22297233
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Somatic hypermutation of the AID transgene in B cells and non-B
          cells
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Direct Submission
JOURNAL   Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
          Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source     1..596
            Location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Ramos 9"
            /product="AID"
            /gene="AID"
            /note="integrated into Burkitt's lymphoma cell line Ramos"
            /codon_start=1
            /protein_id="AA095410.1"
            /db_xref="GI:22297234"
            /translation="MDSLMMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL
            DFGYLNRKNGCHVELLRYISDWDLDFGRCYRVTWFTSWSPCYDCARHVDLFRGNP
            NLSLRIFARLYFCEDRKAEPGLRLHAGVQIAIMTFKDYFYCWNTFVENHETFK
            AWEGHLSVRLSRLRLPLLYEVDLDRFAFRTLGL"
BASE COUNT 128 a 163 c 155 g 150 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      8.06e-201      Length:      596
Score:          198.00        Matches:      198
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:     100.00%      Indels:      0
DB:              9           Gaps:      0

US-09-966-880A-8 (1-198) x AF529823 (1-596)
QY      1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
        |||||||
Db      1 ATGGACAGCCATTATGATGACACCGAGGAAGATTCTTTACCAATTCAAAAATGTCGCTGG 60
        |||||||
QY      21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
        |||||||

```

```

|||||
61 GCTAAGGTCGCGGTGAGACCTACCTGTGCTAGCTAGTAGTGAAGAGCGGTGACAGTCTACA 120
        |||||||
QY      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
        |||||||
Db      121 TCCCTTTTCACGTGAGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 180
        |||||||
QY      61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
        |||||||
Db      181 TTCTCCCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGTCACCGCTCACCTGG 240
        |||||||
QY      81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
        |||||||
Db      241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTCTTCGCGAGGG 300
        |||||||
QY      101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
        |||||||
Db      301 AACCCCAACCTCAGTCTGAGGATCTCACCGCGCCCTCTACTCTGTGAGAGCCGCAAG 360
        |||||||
QY      121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
        |||||||
Db      361 CTGAGCCCGAGGGCTGCGGGCTGCACCGCGGGGTGCCAATAGCCATCATGACC 420
        |||||||
QY      141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
        |||||||
Db      421 TTCAAAGATTATTTTACTGCTGGAACTACTTTGTAGAAAACCATGAAGAACTTTCAAA 480
        |||||||
QY      161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
        |||||||
Db      481 GCCTGGGAGGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCCTT 540
        |||||||
QY      181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
        |||||||
Db      541 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTT 594
        |||||||

```

```

RESULT 2
AF529826
LOCUS      Homo sapiens clone Ramos 12 AID (AID) mRNA, partial cds.
DEFINITION
ACCESSION AF529826
VERSION    AF529826.1 GI:22297239
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Somatic hypermutation of the AID transgene in B cells and non-B
          cells
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS   Martin,A. and Scharff,M.D.
TITLE     Direct Submission
JOURNAL   Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
          Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source     1..596
            Location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Ramos 12"
            /product="AID"
            /gene="AID"
            /note="integrated into Burkitt's lymphoma cell line Ramos"
            /codon_start=1
            /protein_id="AA095413.1"
            /db_xref="GI:22297240"
            /translation="MDSLMMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL
            DFGYLNRKNGCHVELLRYISDWDLDFGRCYRVTWFTSWSPCYDCARHVDLFRGNP
            NLSLRIFARLYFCEDRKAEPGLRLHAGVQIAIMTFKDYFYCWNTFVENHETFK
            NLSLRIFARLYFCEDRKAEPGLRLHAGVQIAIMTFKDYFYCWNTFVENHETFK

```

```
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source
1..596
/organism.."Mus musculus"
/db_xref.."taxon:10090"
/clone.."10"
/cell_type.."hybridoma Pl-5"
/transgenic
1..596
/organism.."Homo sapiens"
/db_xref.."taxon:9606"
1..>596
/gene.."AID"
1..>596
/gene.."AID"
/codon_start..1
/product.."AID"
/protein_id.."AAM95424.1"
/db_xref.."GI:22297262"
/translacion.."MDSLLMNRKFLYQFNKRWAKGRRETYL..YVVKRRDSATSFSL
DFGLRNKNGCHVLELLFLRYISDWDLDPGRCYRVYTWFTSPSCYDCARHVAADFLRGNP
NLSKRIFTARLYFCEDRKAEPGLRRLHRAVGQIAIMTFKDYFCWNTFVENHETFK
AMEGLHNSVRLSRQLRILLPLYEVDDLDAFRITLGL"
BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN
Alignment Scores:
Pred. No.: 8,06e-201 Length: 596
Score: 198.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0
US-09-966-880A-8 (1-198) x AF529837 (1-596)
QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
DB 1 ATGGACACGCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGG 60
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
DB 61 GCTAAGGTCGCGCTGAGACCTTACCTGTGCTACGTAGTGAAGAGCGTCACAGTCTACA 120
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60
DB 121 TCCCTTTTCACTGGACCTTTGGTTATCTTCGCAATAAGAACGCGTCGCCAGATTGGCTC 180
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
DB 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGGTACCTGG 240
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
DB 241 TTCACCTCTCGAGCGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGAGGG 300
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
DB 301 AATCCCAACCTCAGTCTGAGGATCTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
DB 361 GCTGAGCCGAGGGCTCGGCGGCTGCACCGCGCGGGTGCAANTAGCCATCATGACC 420
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
DB 421 TTCAAAGATATATTTTACTGCTGGAAATCTTTTGTAGAAAACCATGAAGAACTTTCAA 480
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
DB 481 GCCTGGGAAGGCTGCATGAANAATTCAGTTCTGCTCTCCAGACAGCTTCGGCGCATCCT 540
QY 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
DB 541 TTGCCCTGTATGAGTTGATGACTTACGAGCGCATTTCTGTACTTTGGGACTT 594
RESULT 3
AF529837
LOCUS Mus musculus clone 10 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529837
VERSION AF529837.1 GI:22297261
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
```

QY 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
 |||||||
 Db 541 TTGCCCTGTATGAGTTGATGACTTACGAGACGCATTCGTACTTTGGGACTT 594

RESULT 4
 BC006296
 LOCUS 1837 bp mRNA linear PRI 12-JUL-2001
 DEFINITION Homo sapiens, activation-induced cytidine deaminase, clone
 MGC:12911 IMAGE:4054915, mRNA, complete cds.
 ACCESSION BC006296
 VERSION BC006296.1 GI:13623400
 KEYWORDS MGC.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1837)
 Strausberg, R.
 Direct Submission
 Submitted (09-APR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-re@mail.nih.gov
 Tissue Procurement: Louis Staudt
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nhgri.nih.gov
 Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
 Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
 Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Snyder, B., Stantropop, S., Thomas, P.J.,
 Tongson, E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
 Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 17 Row: A Column: 1
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 9988409.
 Location/Qualifiers
 1..1837
 /organism="Homo sapiens"
 /db_xref="LocusID:57379"
 /db_xref="taxon:9606"
 /clone="MGC:12911 IMAGE:4054915"
 /tissue-type="Primary B-Cells from Tonsils"
 /clone_lib="NIH_MGC_48"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 86..682
 /codon_start=1
 /product="activation-induced cytidine deaminase"
 /protein_id="AAH06296.1"
 /db_xref="GI:13623401"
 /translation="MDSLMMNRKFLYQFNKRWAKGRRETYLCYVVKRRDSATSFSL
 DFLYRNKNGCVLLFLRYISDWLDPGRCYRVTWFTWSPCYDCARHVADELGRNP
 NLSLRIFFARLYFCEDKAEPEGLRLHAGVQIALMTFKDYFCWNTFVENHETFK
 AWEGLHNSVRLSRQLRLLPLLYEVDLDRDAFTLGL"
 530 a 387 c 421 g 499 t

CDS
 /product="activation-induced cytidine deaminase"
 /protein_id="AAH06296.1"
 /db_xref="GI:13623401"
 /translation="MDSLMMNRKFLYQFNKRWAKGRRETYLCYVVKRRDSATSFSL
 DFLYRNKNGCVLLFLRYISDWLDPGRCYRVTWFTWSPCYDCARHVADELGRNP
 NLSLRIFFARLYFCEDKAEPEGLRLHAGVQIALMTFKDYFCWNTFVENHETFK
 AWEGLHNSVRLSRQLRLLPLLYEVDLDRDAFTLGL"
 530 a 387 c 421 g 499 t

BASE COUNT
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,136-200 Length: 1837
 Score: 198.00 Matches: 198
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x BC006296 (1-1837)

QY 1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
 |||||||
 Db 86 ATGGACAGCCCTCTTGTATGAACGGAGGAAGTTCTTTACCAATTCAAAAATGTCCGCTGG 145

QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
 |||||||
 Db 146 GCTAAGGTCGGCGTGAGACCTACCTGTGCTACGTAGTAGAAGAGCGGTGACAGTGCTACA 205

QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
 |||||||
 Db 206 TCCCTTTTACATGGACTTTGGTTATCTTCGCAATAGAAGCGCTGCCAGTGAATGTCTC 265

QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
 |||||||
 Db 266 TTCTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGGCTCACCTGG 325

QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
 |||||||
 Db 326 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTTCGAGGG 385

QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
 |||||||
 Db 386 AACCCCAACCTCAGTCTGAGGATCTTACCCGGCGCCTCTACTTCTGTGAGGACCGCAAG 445

QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
 |||||||
 Db 446 GCTGAGCCGAGGGGCTCGCGGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACC 505

QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
 |||||||
 Db 506 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCCATGAAAGAACTTTTCAA 565

QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
 |||||||
 Db 566 GCCTGGGAAGGCTCCATGAAATTCAGTTCTCTCCACACAGCTTCGGGGCATCTT 625

QY 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
 |||||||

Db 626 TTGCCCTGTATGAGTTGATGACTTACGAGACGCATTCGTACTTTGGGACTT 679

RESULT 5
 AB040431
 LOCUS 2791 bp mRNA linear PRI 03-OCT-2000
 DEFINITION Homo sapiens AID mRNA for activation-induced cytidine deaminase,
 complete CDS.
 ACCESSION AB040431
 VERSION AB040431.1 GI:9988409
 KEYWORDS AID: activation-induced cytidine deaminase; Human AID.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (sites)
 Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.
 Isolation, tissue distribution, and chromosomal localization of the
 human activation-induced cytidine deaminase (AID) gene
 Genomics 68 (1), 85-88 (2000)
 20408890
 2 (sites)
 Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O.,
 Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A.,
 Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A. G., Brousse, N.,
 Muramatsu, M., Notarangelo, L. D., Kinoshita, K., Honjo, T., Fischer, A.
 and Durandy, A.
 Activation-induced cytidine deaminase (AID) deficiency causes the
 autosomal recessive form of the Hyper-IgM syndrome (HIGM2)
 Cell 102 (5), 565-575 (2000)

TITLE
 JOURNAL

MEDLINE 20460541
REFERENCE 3 (bases 1 to 2791)
AUTHORS Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine; Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan (E-mail:honjo@mfour.med.kyoto-u.ac.jp, Tel:81-75-753-4371(ex.4371), Fax:81-75-753-4388)

FEATURES

Location/Qualifiers

1..2791

/organism="Homo sapiens"

/db_xref="taxon:9606"

1..2791

/gene="AID"

77..673

/gene="AID"

/codon_start=1

/product="activation-induced cytidine deaminase"

/protein_id="BAB12721.1"

/db_xref="GI:9988410"

/translation="MDSLMMNRKFLYQFNVRKGRRETYLCYVVKRRDSATSFSL

DFGLRNKNGCHVELLFLRYISDMDLDPGRCYRVWTFSWSPCDCAHVAADFLRGNP

NLSRIETARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYFCWNTFVENHETFK

AWEGHENSVRSLRQLRILLPLLYEVDLDRDAFRTLGL"

BASE COUNT 842 a 548 c 625 g 776 t

ORIGIN

Alignment Scores:

Pred. No.: 3,066-200 Length: 2791

Score: 198.00 Matches: 198

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AB040431 (1-2791)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
|||||
Db 77 ATGGACACCTCTTGATCAACGGGAGGAGTTCTTTACCAATTCAAAATGTCGGCTGG 136

Qy 21 AlaLysGlyArgArgGluThrTyrLeuCystTyrValValLysArgArgSerAlaThr 40
|||||
Db 137 GCTAAGGTCGGCGTGAGACCTACCTGTGTACGTAGTGTAGTGAAGAGCGGTGACAGTGCTACA 196

Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
|||||
Db 197 TCCITTTTCCATGGACTTTGGTTATCTTCGCAATAAGAACGGGTGCCAGTGGAAATTTGCTC 256

Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
|||||
Db 257 TTCTCTCCGCTACATCTCGGACCTGGGACCTAGACCTGCGCGCTGCTACCGGCTCACCTGG 316

Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
|||||
Db 317 TTCACCTCTGGAGCCCTGCTAGACATGTGCCGACATGTGGCGGACTTTCTTGGAGGG 376

Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
|||||
Db 377 AACCCCAACCTCAGTCTCAGGATCTTACCGCGGCGCTCTACTCTGTGAGGACCGCGCAAG 436

Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
|||||
Db 437 GCTGAGCCGAGGGGCTCGCGGCTGACCGCGCGGGGTGCCAAATAGCCATCATGACC 496

Qy 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
|||||
Db 497 TTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAATTTTCAA 556

Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
|||||
Db 557 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCTCCAGACAGCTTTCGGCGCATCCTT 616

Qy 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
|||||
Db 617 TTGCCCTGTATGAGTTGATGACTTACGACAGCCGATTTTCGTACTTTGGGACTT 670

RESULT 6

AF529829

LOCUS

DEFINITION Mus musculus clone 2 transgenic Homo sapiens AID (AID) mRNA,
complete cds.

ACCESSION AF529829

VERSION AF529829.1

KEYWORDS

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Location/Qualifiers

1..596

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="2"

/cell_type="hybridoma PI-5"

/transgenic

1..596

/organism="Homo sapiens"

/db_xref="taxon:9606"

1..>596

/gene="AID"

1..>596

/gene="AID"

/product="AID"

/protein_id="AAM95416.1"

/db_xref="GI:22297246"

/translation="MDRLMMNRKFLYQFNVRKGRRETYLCYVVKRRDSATSFSL

DFGLRNKNGCHVELLFLRYISDMDLDPGRCYRVWTFSWSPCDCAHVAADFLRGNP

NLSRIETARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYFCWNTFVENHETFK

AWEGHENSVRSLRQLRILLPLLYEVDLDRDAFRTLGL"

BASE COUNT 128 a 163 c 155 g 150 t

ORIGIN

Alignment Scores:

Pred. No.: 1,246-197 Length: 596

Score: 195.00 Matches: 195

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 98.48% Indels: 0

DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529829 (1-596)

Qy 4 LeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGly 23
|||||
Db 10 CTCTTTGATGACCGGAGGAGTTTCTTTACCAATTCAAAATGTCGGCTGAGCTACGGT 69

Qy 24 ArgArgGluThrTyrLeuCysTyrValValLysArgArgSerAlaThrSerPheSer 43
|||||
Db 70 CGGCGTGAGACCTACCTGTCTAGTACTGAGAGGCGTGACGTGCTACATCTCTTTCA 129

Qy 44 LeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeuPheLeuArg 63
|||||
Db 130 CTGCACTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGCTGGAAATTCCTCTCCGC 189

Qy 64 TyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSer 83

```
Db 190 TACATCTCGGACTGGAGCTAGACCTGGCGCTGCTACCGCGTCACCTGTTCACTCC 249
Qy 84 TrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsn 103
Db 250 TGGAGCCCTGCTAGGACTGTGCCGACATGTGCCGACTTCTGCGAGGAAACCCCAAC 309
Qy 104 LeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgGlyAlaGluPro 123
Db 310 CTCAGTCTGAGGATCTTCACCGCGCTCTACTTCTGTGAGGACCGCAAGCTGAGGCC 369
Qy 124 GluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAsp 143
Db 370 GAGGGCTGCGCGCTGCACCGCGCGGTGCAAAATAGGCATCATGACCTTCAAGAT 429
Qy 144 TyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGlu 163
Db 430 TATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAAGCTTCAAGACCTGGAA 489
Qy 164 GlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLeuProLeu 183
Db 490 GGGCTGCATGAAATTCAGTTGCTCTCCACAGAGCTTCGCGGCATCTTTGCGCCGTG 549
Qy 184 TyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 550 TATGAGGTTGATGACTTACGAGAGCGCAATTCGTACTTTGGGACTT 594

RESULT 7
AF529842 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 2 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529842
VERSION AF529842.1 GI:22297271
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="2"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AA095429.1"
/db_xref="GI:22297272"
/translation="MDSLMMRRKFLYQKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLEFLRYISDWLDPGRCYRVTFWTSNPSPCDARHVAFLRGNP
NLSLRIFARLYFCEDRKAEPGLRLRHAGVOIAIMTFDYFCWNTFVFNHRTFK
AWEGHNSVRLSRQLRILLPLYEVDLDAFTWGR"

BASE COUNT 127 a 164 c 157 g 149 t
```

```
ORIGIN
Alignment Scores: 1.25e-197 Length: 597
Pred. No.: 195.00 Matches: 195
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 98.48%
DB: 12 Gaps: 0
US-09-966-880a-8 (1-198) x AF529842 (1-597)
Qy 1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGACAGCCCTCTTATGACCCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCGTGG 60
Qy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTCTACA 120
Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATGAAGACGGCTGCCACGTGGAATTGCTC 180
Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCCTTCGCTACATCTCGGACTTCGGACCTAGACCCCTGCGCGCTGCTACCGCTCACCTGG 240
Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTCTGGAGCCCTGCTACGACTGTCCCGACATGTGCGCGACTTTCTGCGAGGG 300
Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCAACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGGTGCMAAATAGCCATCATGACC 420
Qy 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAGAGATATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAAGACTTTCAA 480
Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 481 GCCTGGGAAGGCGTGCATGAAATTCAGTCTCTCTCCACAGACGCTTCGGCGCATCCTT 540
Qy 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThr 195
Db 541 TTGCCCTGTATGAGGTTGATGACTTACGTGACGCGCATTTTCGTACT 585

RESULT 8
AF529843 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 3 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529843
VERSION AF529843.1 GI:22297273
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
```

JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES
source 1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="3"
/cell_line="CHO"
/transgenic
source 1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene 1..597
/gene="AID"
CDS 1..597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95430.1"
/db_xref="GI:22297274"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRDSATSFSL
DFGLRNKNGCHVELLEFLRYISDWDLDPGRCYRVWTSMSPCYDCARHVDPLRGNP
NLSRIPTARLYFCEDRKAEPEGLRLRHAGVOIAITMFKDYFYCMWTFVENHERTFK
AWEGLHNSVRLSQLRILLPLYEVDLDRDAFTWGR"
BASE COUNT 129 a 164 c 156 g 148 t
ORIGIN

Alignment Scores:
Pred. No.: 1.25e-197 Length: 597
Score: 195.00 Matches: 195
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.48% Indels: 0
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529843 (1-597)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCGTGG 60

Qy 21 AlalysGlyArgArgGluThrTyrLeuCystTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAGGGTCGGCTGAGACCTACCTGCTGCTAGCTAGTGAAGAGCGCTGACAGTCTACA 120

Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCNATGAAGACGCTCCACGTGGAAATTGCTC 180

Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGCGTCAACCTGG 240

Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTGGAGCCCTGCTAGACTGTGCGCGACATGTGGCGGACTTTCTCGGAGGG 300

Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACC CGCGCGCTCTACTTCTGTGAGGACCGCAAG 360

Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 420

Qy 141 PheLysAspTyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAAGATTATTTTACTGCTGGAATACCTTTGTAGAAAAACCATGAAGAACTTTCAAA 480

Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 481 GCCTGGGAAGGGCTACATGAATAATTCATGCTCTCTCCAGACAGCTTCCTGGCGCATCCT 540

Qy 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThr 195
Db 541 TTGCCCCCTGATGAGGTTGATGACTTAGGAGACGCATTTCTGTACT 585

RESULT 9
AF529844 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 4 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529844
VERSION AF529844.1 GI:22297275
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct A Substitution
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source 1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="4"
/cell_line="CHO"
/transgenic
source 1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene 1..597
/gene="AID"
CDS 1..597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95431.1"
/db_xref="GI:22297276"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRDSATSFSL
DFGLRNKNGCHVELLEFLRYISDWDLDPGRCYRVWTSMSPCYDCARHVDPLRGNP
NLSRIPTARLYFCEDRKAEPEGLRLRHAGVOIAITMFKDYFYCMWTFVENHERTFK
AWEGLHNSVRLSQLRILLPLYEVDLDRDAFTWGR"
BASE COUNT 128 a 163 c 158 g 148 t
ORIGIN

Alignment Scores:
Pred. No.: 1.25e-197 Length: 597
Score: 195.00 Matches: 195
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.48% Indels: 0
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529844 (1-597)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCGTGG 60

Qy 21 AlalysGlyArgArgGluThrTyrLeuCystTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAGGGTCGGCTGAGACCTACCTGCTGCTAGCTAGTGAAGAGCGCTGACAGTCTACA 120

Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCNATGAAGACGCTCCACGTGGAAATTGCTC 180

```
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
|||||
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGCTCACCTGG 240

QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
|||||
Db 241 TTCACTCTCTGGAGCGCCCTGCTACGACTGTGCCGACATGTGCCGACTTCTTGGGAGG 300

QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
|||||
Db 301 AACCCCAACCTCAGTCTGAGGATCTACCGCGCGGCTCTACTTCTGTGAGGACCGGAG 360

QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
|||||
Db 361 GCTGAGCCGAGGGCTGGCGGCTGCACCGCGGGGTGCAAAATAGCCATCATGACC 420

QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
|||||
Db 421 TTCAAGATATATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAAGAACTTTCAA 480

QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgIleLeu 180
|||||
Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGACACGCTTCCGGCGATCTT 540

QY 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThr 195
|||||
Db 541 TTGCCCTCTATGAGGTTGATGACTTACGAGACGCAATTCGTACT 585

RESULT 10
AF529847 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 7 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529847
VERSION AF529847.1 GI:22297281
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Cricetulus griseus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source 1..597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="7"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AA095434.1"
/db_xref="GI:22297282"
/translation="WDSLIMNRKFLYQFNVRWAKGRRETYLCYVVKRDSATSFSL
DFGYLRNKGCHVELFLRYISDWLDPRGYRVTWFTSWSPCYCARHVDFLRGNP
NLSLRIFTARLFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNFTVENHRTFK
```

```
BASE COUNT 129 a 164 c 156 g 148 t
ORIGIN
AWEGLEHNSVRLSRQLRRILLPLYEVDDLDRDAFTWGR*

Alignment Scores: 1.25e-197 Length: 597
Pred. No.: 195.00 Matches: 195
Score: 100.00 Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 98.48% Gaps: 0
DB: 12
US-09-966-880A-8 (1-198) x AF529847 (1-597)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTTrp 20
|||||
Db 1 ATGGACAGCCCTCTTGATGACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCCGCTGG 60

QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
|||||
Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTAGTGAGAGGCGTGACAGTGCTACA 120

QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
|||||
Db 121 TCCTTTTTCACCTGGAGCTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCCTC 180

QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
|||||
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTTGGCGGCTGTACCGCTCACCTGG 240

QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
|||||
Db 241 TTCACTCTCTGGAGCGCCCTGCTACGACTGTGCCGACATGTGCCGACTTCTTTCGCGAGG 300

QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
|||||
Db 301 AACCCCAACCTCAGTCTGAGGATCTACCGCGCGGCTCTACTTCTGTGAGGACCGCAAG 360

QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
|||||
Db 361 GCTGAGCCGAGGGCTGGCGGCTGCACCGCGGGGTGCAAAATAGCCATCATGACC 420

QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
|||||
Db 421 TTCAAGATATATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAAGAACTTTCAA 480

QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgIleLeu 180
|||||
Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTCGCTCTCCAGACACGCTTCCGGCGATCTT 540

QY 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThr 195
|||||
Db 541 TTGCCCTCTATGAGGTTGATGACTTACGAGACGCAATTCGTACT 585

RESULT 11
AF529852 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 12 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529852
VERSION AF529852.1 GI:22297291
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Cricetulus griseus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
```

AUTHORS Martin.A. and Scharff,M.D.
 TITLE Direct Submission
 JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES
 source
 1..597
 /organism="Cricetulus griseus"
 /db_xref="taxon:10029"
 /clone="12"
 /cell_line="CHO"
 /transgenic
 1..597
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..597
 /gene="AID"
 1..597
 /gene="AID"
 1..597
 /codon_start=1
 /product="AID"
 /protein_id="AAM95439.1"
 /db_xref="GI:2297292"
 /translation="MDSLNNRRKFLYQKKNVRWAKGRRETYLCYVVKRRDSATSFSL
 DFGYLRNKGCHVELLRYISDWLDPGRCYRVTFWTSWSPCYDCARHVAFLRGNP
 NLSLRIFTARLYFCEDRKAEPGLRLHAGVQIAIMTFKDYFCWNTFVENHRTFK
 AWEGLHENSRLSRLRILLPLYEVDDLRFDAFTWGR"

BASE COUNT 129 a 164 c 156 g 148 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,25e-197 Length: 597
 Score: 195.00 Matches: 195
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 98.48% Indels: 0
 DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529852 (1-597)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
 Db 1 ATGGACAGCCTTTGATGAACGGAGGAAGTTCTTTACCAATTCAAAATTCGGCTGG 60
 QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
 Db 61 GCTAAGGTCGGCGTGAGACCTACCTGTCTAGTAGTGAAGCGGTGACAGTGCTACA 120
 QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
 Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGAATAAGACGGCTGCCACGTGCAATTGCTC 180
 QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
 Db 181 TTCTCCCGCTACATCTCGGACTGGGACCTAGACCTGGCCCTGGCTACCGCTACCTGG 240
 QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
 Db 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGCACTTCTGCGAGGG 300
 QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
 Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
 QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
 Db 361 GCTGAGCCCGAGGAGCTGGCGCGCTGGCACCGCGGGGTGCAATAGCCATCATGACC 420
 QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
 Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAAACCATGAAGAAGACTTTCAAA 480
 QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180

Db 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 540
 QY 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThr 195
 Db 541 TTGCCCCCTGTAGAGGTTGATGACTTACGAGAGCGCATTTTCGTACT 585

RESULT 12
 AF529846
 LOCUS Cricetulus griseus clone 5 transgenic Homo sapiens AID (AID) mRNA, complete cds. 597 bp mRNA linear SYN 19-AUG-2002
 ACCESSION AF529846
 VERSION AF529846
 KEYWORDS GI:2297292
 SOURCE Chinese hamster.
 ORGANISM Cricetulus griseus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 Cricetulus.
 REFERENCE 1 (bases 1 to 597)
 AUTHORS Martin,A. and Scharff,M.D.
 TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 597)
 AUTHORS Martin,A. and Scharff,M.D.
 TITLE Direct Submission
 JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES
 source
 1..597
 /organism="Cricetulus griseus"
 /db_xref="taxon:10029"
 /clone="16"
 /cell_line="CHO"
 /transgenic
 1..597
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..597
 /gene="AID"
 1..597
 /gene="AID"
 /codon_start=1
 /product="AID"
 /protein_id="AAM95433.1"
 /db_xref="GI:2297280"
 /translation="MDSLNNRRKFLYQKKNVRWAKGRRETYLCYVVKRRDSATSFSL
 DFGYLRNKGCHVELLRYISDWLDPGRCYRVTFWTSWSPCYDCARHVAFLRGNP
 NLSLRIFTARLYFCEDRKAEPGLRLHAGVQIAIMTFKDYFCWNTFVENHRTFK
 AWEGLHENSRLSRLRILLPLYEVDDLRFDAFTWGR"

BASE COUNT 128 a 165 c 157 g 147 t
 ORIGIN

Alignment Scores:
 Pred. No.: 2,57e-192 Length: 597
 Score: 190.00 Matches: 190
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 95.96% Indels: 0
 DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529846 (1-597)

QY 6 MetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgArg 25
 Db 16 ATGAACCGAGGAAGTTCTTTACCAATTCAAAATTCCTGGCTAAGCGTCGCGT 75
 QY 26 GluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPheSerLeuAsp 45
 Db 76 GAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACGTGCTACATCTTTTACATGGAC 135
 QY 46 PheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuPheLeuArgIle 65


```
Db 136 TTTGGTTATCTCGCAATAAGACGGCTGCCACGTGGAATTGCTTCTCCTCCGCTACATC 195
Qy 66 SerAspTtpAspLeuAspProGlyArgCysTyrArgValThrTtpPheThrSerTtpSer 85
Db 196 TCGACTGGACCTAGACCCCTGGCGCTGCTACCGCTACCTGTTCCACCTCTCTGGAGC 255
Qy 86 ProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeuSer 105
Db 256 CCCTGCTACGACTGTGCCGACATGTGCGCGACTTCTTCTGCGAGGGAACCCCAACCTCAGT 315
Qy 106 LeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGluGly 125
Db 316 CTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGG 375
Qy 126 LeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPhe 145
Db 376 CTGGCGCGCTGCACCGCGCGGGGTGCAATAGCCATCATGACCTTCAAAGATATTTT 435
Qy 146 TyrCysTtpAsnThrPheValIgluAsnHisGluArgThrPheLysAlaTtpGluGlyLeu 165
Db 436 TACTGCTGGAATCTTTTGTAGAAAACCATGAAGAACTTCAAAGCCTGGGAAGGGCTG 495
Qy 166 HisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLeuProLeuTyrGlu 185
Db 496 CATGAAAATTCAGTTCGCTCTCAGACAGCTTCGGCGCATCTTTTGGCCCTGTATGAG 555
Qy 186 ValAspAspLeuArgAspAlaPheArgThr 195
Db 556 GTTGATGACTTACGAGACCCATTTTCGTACT 585

RESULT 13
AF529832 596 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 5 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529832
VERSION AF529832.1 GI:22297251
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 596)
JOURNAL Martin,A. and Scharff,M.D.
AUTHORS Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 596)
TITLE Martin,A. and Scharff,M.D.
JOURNAL Direct Submission
AUTHORS Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
JOURNAL Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source 1..596
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="5"
/cell_type="hybridoma P1-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAMP95419.1"
/db_xref="GI:22297252"
/translation="MDSLMMRRKFLYOFKNVRWAKGPRRETYLCYVVKRRDSATSESL
DFGYLRNKGCHVELLFRLYSIDNDLDFGRCYRVWTWSNPSFCDCAKHVADFLRGNP
NLSLRIFTARLYFCEDRKAPEGLRLRHAGVQIAIMTFKDYFCWNTFVENHERTEK
```

```
BASE COUNT 129 a 164 c 153 g 150 t
ORIGIN
AWEGLHENSRLSRQLRHILLPLYEVDLDRDAFRTLGL"
Alignment Scores: 1.68e-178 Length: 596
Pred. No.: 177.00 Matches: 177
Score: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 89.39% Indels: 0
DB: 12 Gaps: 0
US-09-966-880A-8 (1-198) x AF529832 (1-596)
Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTtp 20
Db 1 ATGCACAGCCCTTTGATGAAACCGGAGGAAGTTCTTTTACCANTTCAAAAATGTCGCTGG 60
Qy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAAGGTGCGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGCTACA 120
Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValIgluLeuLeu 60
Db 121 TCCCTTTTCACTGGAGCTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTC 180
Qy 61 PheLeuArgTyrIleSerAspTtpAspLeuAspProGlyArgCysTyrArgValThrTtp 80
Db 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCCGCTGCTACCGGCTCACCT 240
Qy 81 PheThrSerTtpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCCGAGCTTTCTGCGAGGG 300
Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCCGCGCGCTCTACTTCTGTGAGAGCCGCAAG 360
Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCCGAGGGGCTCGCGCGGCTGCACCCGCGCGGGTGCATAATAGCCATCATGACC 420
Qy 141 PheLysAspTyrPheTyrCysTtpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAAGATATTTTACTGCTGGAATCTTTGTAGAAAACCATGAAGAACTTCAAA 480
Qy 161 AlaTtpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArg 177
Db 481 GCCTGGGAAGGGCTGTCATGAAAATTCAGTTCGCTCTCCACAGACGTTTCGG 531
```

RESULT 14

AF529815

LOCUS

Homo sapiens clone Ramos 1 AID (AID) mRNA, partial cds.

DEFINITION

AF529815

ACCESSION

AF529815.1

VERSION

GI:22297217

KEYWORDS

human.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 1"
1..>596
/gene="AID"
1..>596
/gene="AID"
/notes="integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95402.1"
/db_xref="GI:22297218"
/translacion="MDSLLMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKNCHVELLEFLRYISDMWDPGRCYRVTFWTSBSPCYDCAHVADELGRNP
NLSLRIFATRLYFCEDRKAEPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHERTFK
AMEGLHENSRLSRQIRILLPLVEVDLDRFRTLGL"
BASE COUNT 128 a 163 c 155 g 150 t
ORIGIN

Alignment Scores:
Pred. No.: 2,24e-176 Length: 596
Score: 175.00 Matches: 175
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 88.38% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AF529815 (1-596)
Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGCACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
Qy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAGGTCGGCTGAGACCTACTGTCTAGCTAGTGAAGAGCGCGTACAGTGTCTACA 120
Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATGAAGACGCTCCACGTTGGAATTCGTC 180
Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCGCGTCACTGG 240
Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTGGAGCCCTGCTAGACTGTGCCGACATGTGCCGACTTTCTCGGAGGG 300
Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGCTGAGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGTGCAAAATAGCATCATGACC 420
Qy 141 PheLysAspTyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAGATATATTTTACTGCTGGAATACATTTGTAGAAAACCAATGAAGAACTTCCAA 480
Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175
Db 481 GCCTGGGAAGGCTGCATGAATAATTCAGTTCGTCTCTCCACAGAG 525

RESULT 15
AF529831
LOCUS
DEFINITION Mus musculus clone 4 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
ACCESSION AF529831
VERSION AF529831.1 GI:22297249
KEYWORDS
```

```
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 596)
TITLE Martin, A. and Scharff, M.D.
JOURNAL Somatic hypermutation of the AID transgene in B and non-B cells
Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="4"
/cell_type="hybridoma Pl-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95418.1"
/db_xref="GI:22297250"
/translacion="MDSLLMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKNCHVELLEFLRYISDMWDPGRCYRVTFWTSBSPCYDCAHVADELGRNP
NLSLRIFATRLYFCEDRKAEPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHERTFK
AMEGLHENSRLSRQIRILLPLVEVDLDRFRTLGL"
BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN

Alignment Scores:
Pred. No.: 2,24e-176 Length: 596
Score: 175.00 Matches: 175
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 88.38% Indels: 0
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529831 (1-596)
Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGCACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
Qy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAGGTCGGCTGAGACCTACTGTCTAGCTAGTGAAGAGCGCGTACAGTGTCTACA 120
Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATGAAGACGCTCCACGTTGGAATTCGTC 180
Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCGCGTCACTGG 240
Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTGGAGCCCTGCTAGACTGTGCCGACATGTGCCGACTTTCTCGGAGGG 300
Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGCTGAGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGTGCAAAATAGCATCATGACC 420
Qy 141 PheLysAspTyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAGATATATTTTACTGCTGGAATACATTTGTAGAAAACCAATGAAGAACTTCCAA 480
Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175
Db 481 GCCTGGGAAGGCTGCATGAATAATTCAGTTCGTCTCTCCACAGAG 525

RESULT 15
AF529831
LOCUS
DEFINITION Mus musculus clone 4 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
ACCESSION AF529831
VERSION AF529831.1 GI:22297249
KEYWORDS
```

```
Db 361 GCTAGCCGAGGGCTCGCGCGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACC 420
|||||
Qy 141 PheLysAspTyrPheTyrCysTyrPheValGluAsnHisGluArgThrPheLys 160
|||||
Db 421 TTCAAGATTATTTTACTGCTGGATACTTTTGTAGAAACCATGAAAGAACTTTCANA 480
|||||
Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175
|||||
Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCTCCAGACAG 525
|||||

RESULT 16
AF529851 597 bp mRNA linear SYN 19-AUG-2002
LOCUS Cricetulus griseus clone 11 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529851
VERSION AF529851.1 GI:22297289
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus.
REFERENCE 1 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1. 597
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="11"
/cell_line="CHO"
/transgenic
2. 597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/gene="AID"
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95438.1"
/db_xref="GI:22297290"
/translation="MDSLMMNRKFLYQFKNVRWAKGRREYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELLRLYISDMDLPGRCYRTWTSPCYDCARHVAFLRGNP
NLSRITARLYFCEDRKAEPEGLRLRHAGVQIAIMTFKDYFCWNTFVENHERTFK
AWEGHNSVRLSRQLRILLPLVEVDLDRDAFRTWGR"
BASE COUNT 128 a 165 c 157 g 147 t
ORIGIN

Alignment Scores:
Pred. No.: 3.46e-173 Length: 597
Score: 172.00 Matches: 172
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 86.87% Indels: 0
Db: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529851 (1-597)

Qy 1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
|||||
Db 1 ATGGACACCTCTTGATGACCGGGAAGGTTCTTTACCAATTCAAAAATGTCGCGTGG 60
|||||
```

```
Qy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
|||||
Db 61 GCTAAGGGTCGCGTGCAGACCTACCTGTGTACGTAGTGAAGAGCGGTGACAGTGTACA 120
|||||
Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60
|||||
Db 121 TCCTTTTCTACGTGGACTTTGTTATCTTCGCAATAAGAACGGCTGCCAGCTGGAATTGCTC 180
|||||
Qy 61 PheLeuArgTyrIleSerAspTyrAspLeuAspProGlyArgCysTyrArgValThrTrp 80
|||||
Db 181 TTCCTCCGCTACATCTCGGACTGGGACTAGACCTGGCCGCTGCTACCGCGTCACTGG 240
|||||
Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
|||||
Db 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTTGGCAGGG 300
|||||
Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
|||||
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCACCGCGCCTTACTTCTGTGAGGACCGCAAG 360
|||||
Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
|||||
Db 361 GCTGAGCCGAGGGCTGCGCGGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACC 420
|||||
Qy 141 PheLysAspTyrPheTyrCysTyrPheValGluAsnHisGluArgThrPheLys 160
|||||
Db 421 TTCAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCANA 480
|||||
Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeu 172
|||||
Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTCTGCTCTC 516
|||||

RESULT 17
AF529825 596 bp mRNA linear PRI 19-AUG-2002
LOCUS Homo sapiens clone Ramos 11 AID (AID) mRNA, partial cds.
DEFINITION AF529825
ACCESSION AF529825.1 GI:22297237
VERSION AF529825.1
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1. 596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 11"
2. >596
/gene="AID"
3. >596
/gene="AID"
/notes="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95412.1"
/db_xref="GI:22297238"
/translation="MDSLMMNRKFLYQFKNVRWAKGRREYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELLRLYISDMDLPGRCYRTWTSPCYDCARHVAFLRGNP
NLSRITARLYFCEDRKAEPEGLRLRHAGVQIAIMTFKDYFCWNTFVENHERTFK
AWEGHNSVRLSRQLRILLPLVEVDLDRDAFRITGL"
BASE COUNT 126 a 164 c 156 g 150 t
ORIGIN
```



```

ACCESSION AF529833
VERSION AF529833.1 GI:22297253
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Martin.A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE
AUTHORS Martin.A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="6"
/cell_type="hybridoma P1-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95420.1"
/db_xref="GI:22297254"
/translocation="MDSLLMNRKELYOFKNVRWAKGRRETYVCVVKRRDSATSFSL
DFGYLRNKGCHVELLFURYISDWLDGRCYRVTFWTSWSPCYDCARHVAFLRGNP
NLSRIFTARLYFCEDRAEPGLRLRHAGVOIAIMTFKDYFCWNTFVENHERTEFK
AWEGUHSVRLSRQLRLLPLYEVDLDRDAFTLGL"
BASE COUNT 127 a 163 c 156 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 5,32e-170 Length: 596
Score: 169.00 Matches: 197
Percent Similarity: 98.99% Conservative: 0
Best Local Similarity: 98.99% Mismatches: 1
Query Match: 85.35% Indels: 2
DB: 12 Gaps: 0
US-09-966-880A-8 (1-198) x AF529833 (1-596)
QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTtr 20
Db 1 ATGGACAGCCTCTTGATGAAACGGAGAGATTCTTTACCAATTCAAAATGTCGCTGG 60
QY 21 AlaLysGlyArgArgGluThrTyrLeu-CystTyrValValLysArgArgAspSerAlath 40
Db 61 GCTAAGGTCGCGCTGAGACCTACGT-GRGCTACGTAGTGAAGAGCGGTGACAGTGCTAC 119
QY 40 rSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLe 60
Db 120 ATCCCTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGAATTGCT 179
QY 60 uPheLeuArgTyrIleSerAspTtrAspLeuAspProGlyArgCysTyrArgValThrTr 80
Db 180 CTTCTCCGCTACACTCGACTCGGACCTAGACCTGCGCCGCTGCTACCGGCTCACCTG 239
QY 80 pPheThrSerTtrPserProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGl 100
Db 240 GTTCACCTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTCGGAGG 299
QY 100 yAsnProAnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLy 120
|||||

```

```

Db 300 GAACCCCAACCTCAGTCTCAGGATCTTCACCCGCGCCTCTACTTCTGTGAGACCGCAA 359
QY 120 sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
|||||
Db 360 GGCTAGCCCGAGGGGCTGCGCGGCTGCACCCGCGGGGTGCAAAATAGCCATCATGAC 419
QY 140 rPheLysAspTyrPheTyrCysTtrPAsnThrPheValGluAsnHisGluArgThrPheLy 160
|||||
Db 420 CTTCAAGATATTATTTTACTGCTGGAATACTTTTGTAGAAAAACCATGAAGAATTTCAA 479
QY 160 sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
|||||
Db 480 ACCCTGGGAAGGCTGCATGAAATTCAGTTCGTCTCTCCACAGACGTTCCGCGCATCCT 539
QY 180 uLeuProLeuTyrGluValAspLeuArgAspLeuArgAspPheArgThrLeuGlyLeu 198
|||||
Db 540 TTTGGCCCTGTATGAGTTGATGACTTACGAGACCGCATTTCTACTTTGGGACTT 594

RESULT 20
AF529835
LOCUS Mus musculus clone 8 transgenic Homo sapiens AID (AID) mRNA, SYN 19-AUG-2002
DEFINITION complete cds.
ACCESSION AF529835
VERSION AF529835.1 GI:22297257
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Martin.A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE
AUTHORS Martin.A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Source
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="8"
/cell_type="hybridoma P1-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95422.1"
/db_xref="GI:22297258"
/translocation="MDSLLMNRKELYOFKNVRWAKGRRETYVCVVKRRDSATSFSL
DFGYLRNKGCHVELLFURYISDWLDGRCYRVTFWTSWSPCYDCARHVAFLRGNP
NLSRIFTARLYFCEDRAEPGLRLRHAGVOIAIMTFKDYFCWNTFVENHERTEFK
AWEGUHSVRLSRQLRLLPLYEVDLDRDAFTLGL"
BASE COUNT 127 a 163 c 156 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 5,32e-170 Length: 596
Score: 169.00 Matches: 197
Percent Similarity: 98.99% Conservative: 0
Best Local Similarity: 98.99% Mismatches: 1
Query Match: 85.35% Indels: 2
DB: 12 Gaps: 0
US-09-966-880A-8 (1-198) x AF529835 (1-596)

```

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTTP 20
 Db 1 ATGGACAGCCTCTGATGAACCGGAGGAGTTCTTTACCAATTCAAAAATGTCGCGTGG 60
 QY 21 AlaLysGlyArgArgGluThrTyrLeu-CysTyrValValLysArgArgSerAlaTh 40
 Db 61 GCTAAGGGTCGGCGTGAGACCTACGT-GTGCTACGTAGTGAAGAGCGGTGACAGTGCTAC 119
 QY 40 rSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLe 60
 Db 120 ATCCCTTTTACGGGACTTTGGTTATCTTCGCAATAAGAACGCGCTGCCAGTGGAAATGCT 179
 QY 60 uPheLeuArgTyrIleSerAspTyrAspLeuAspProGlyArgCysTyrArgValThrTr 80
 Db 180 CTTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCGGCTCACCTG 239
 QY 80 pPheThrSerTyrSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGl 100
 Db 240 GTTCACCTCTCGAGCCCTGCTAGGACTGTGCCGACATGTGGCGGACTTTCTCGGAGG 299
 QY 100 vAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgly 120
 Db 300 GAACCCACCTCAGTCTGAGGATCTTACCCGCGGCTCTACTTCTGTGAGGACCGCAA 359
 QY 120 salAGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
 Db 360 GGTGAGCCCGAGGGCTGCGCGGCTGCACCCGCGCGGGTGCAATAGCCATCATGAC 419
 QY 140 rPheLysAspTyrPheTyrCysTyrPAsnThrPheValGluAsnHisGluArgThrPheLy 160
 Db 420 CTTCAAGATTATTTTACTGTGGAATACTTTGTAGAAAACCATGAAAGAACTTTCAA 479
 QY 160 salATrPGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
 Db 480 ACCCTGGGAGGCGTGCAATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGGATCCT 539
 QY 180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
 Db 540 TTTGCCCTGTATGAGTTGATGACTTACGAGACGACATTCGTACTTTGGGACTT 594
 RESULT 21
 AF529856 547 bp mRNA linear SYN 19-AUG-2002
 LOCUS Cricetulus griseus clone 16 transgenic Homo sapiens AID (AID) mRNA, complete cds.
 DEFINITION
 ACCESSION AF529856
 VERSION
 KEYWORDS
 SOURCE Chinese hamster.
 ORGANISM Cricetulus griseus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus
 REFERENCE 1 (bases 1 to 547)
 AUTHORS Martin.A. and Scharff,M.D.
 TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 547)
 AUTHORS Martin.A. and Scharff,M.D.
 TITLE Direct Submission
 JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
 FEATURES
 Location/Qualifiers
 source 1..547
 /organism="Cricetulus griseus"
 /db_xref="taxon:10029"
 /clone="16"
 /cell_line="CHO"
 /transgenic
 source 1..547
 /organism="Homo sapiens"

/db_xref="taxon:9606"
 <1..547
 /gene="AID"
 CDS <1..547
 /gene="AID"
 /codon_start=2
 /product="AID"
 /protein_id="AAM95443.1"
 /db_xref="GI:22297300"
 /translation="VRWAKGRRETYLCYVVKRRDSATSFSLDFYLRKNKGCHVELLP
 LRYISDWLDGRCYRVYTWFTSWPCYDCARHVADF LRGNPNLSLRITFARLYFCEDR
 KAEGLRLRHAGVQIAIMTFKDYFYCNWTFEVENHRTFKAEGLHENSRLSRLR
 RIILPLYEVDDLRLDAFTWGR"
 BASE COUNT 112 a 153 c 147 g 135 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.36e-162 Length: 547
 Score: 162.00 Matches: 162
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 81.82% Indels: 0
 DB: 12 Gaps: 0
 US-09-966-880A-8 (1-198) x AF529856 (1-547)
 QY 18 ValArgTTPAlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAsp 17
 Db 2 GTCCCTGGCTTAAGGTCGGCTGAGACCTACCTGCTGCTAGTGAAGAGGCTGAC 61
 QY 38 SerAlaThrSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisVal 57
 Db 62 AGTGTACATCCTTTTCTACCTGACCTTTGCTTATCTTCGCAATAAGAACGGTTCGCCACG 121
 QY 58 GluLeuLeuPheLeuArgTyrIleSerAspTyrPAsnThrPheValGluAsnHisGluArg 77
 Db 122 GAATTCCTCTTCCTCCGCTACATCTCGGACTCGGACCTGAGACCTGCGCGTTCGCTACG 181
 QY 78 ValThrTTPPheThrSerTyrPAsnThrPheValGluAsnHisGluArgHisValAlaAspPhe 97
 Db 182 GTACCTGGTTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCCACTTTT 241
 QY 98 LeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGlu 117
 Db 242 CTCGAGGAGAACCCCAACCTCAGTCTGAGGATCTTCACGCGCGCTCTCTTCTGTGAG 301
 QY 118 AspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAla 137
 Db 302 GACCCGAGGCTGAGCCCGAGGGGCTGCGCGGCTGACCCGCGGGGTGCCAATATAGCC 361
 QY 138 IleMetThrPheLysAspTyrPheTyrCysTyrPAsnThrPheValIleAsnHisGluArg 157
 Db 362 ATCATGACCTTCAAGATTTATTTTACTGCTGGAATACCTTTTGTAGAAAACCATGAAGA 421
 QY 158 ThrPheLysAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArg 177
 Db 422 ACTTTCAAAGCCTGGGAAGGCTGCATGAAATTCAGTTTCGTCCTCTCCACAGACGCTCGG 481
 QY 178 ArgIle 179
 Db 482 CGCATC 487
 RESULT 22
 AF529822 596 bp mRNA linear PBI 19-AUG-2002
 LOCUS Homo sapiens clone Ramos 8 AID (AID) mRNA, partial cds.
 DEFINITION
 ACCESSION AF529822
 VERSION AF529822.1
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 8"
1..>596
/gene="AID"
1..>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AA095409.1"
/db_xref="GI:22297232"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDRATSFSL
DFGLRNKNGCHVELFLRYISDWLDPPGRYRVTFWTSWSPCYDCARHVADEFLRGNP
NLSLRIFTARLYFCEDRKAEPLRLRLHRAVGQIAIMTFKDYFCWNTEFVENHERTFK
AWEGLEHNSVRLSRQLRILLPLYEVDLDRFRTLGL"
BASE COUNT 126 a 165 c 155 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 1.95e-160 Length: 596
Score: 160.00 Matches: 197
Percent Similarity: 98.99% Conservative: 0
Best Local Similarity: 98.99% Mismatches: 1
Query Match: 98.81% Indels: 2
DB: 9 Gaps: 0
US-09-966-880A-8 (1-198) x AF529822 (1-596)
Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
Qy 21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSer-AlaTh 40
Db 61 GCTAAGGGTCGGCTGAGACCTACCTGTGTACGTAGTGAAGAGCGGTGACCG-TGCTAC 119
Qy 40 rSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLe 60
Db 120 ATCTTTTCTACTGGACTTGTGTTATCTTCGCAATAAGAACGGCTGCCACCTGGAATTGCT 179
Qy 60 uPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTr 80
Db 180 CTTCTCTCCGTATACATCTCGAGCTGGGACCTAGACCTGGCGCTGTCTACCGCTCACCTG 239
Qy 80 pPheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGl 100
Db 240 GTTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTCCTCGCAGG 299
Qy 100 yAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLy 120
Db 300 GAACCCCAACCTCAGTGTAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAA 359
Qy 120 sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
Db 360 GGCTGACCCCGAGGGGCTGGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGAC 419
Qy 140 rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160
Db 420 CTTCAAGATATATTTTACTGCTGGAATATCTTTGTAGAAACCATGAAAGAACTTTTCAA 479

Qy 160 sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
Db 480 AGCCTGGGAAGGCGTCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCTT 539
Qy 180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 540 TTTGCCCTCTATGAGGTTGATGACTTACGAGACGACATTCGTACTTTGGGACTT 594
RESULT 23
AF529834
LOCUS Mus musculus clone 7 transgenic Homo sapiens AID (AID) mRNA, 595 bp mRNA linear SYN 19-AUG-2002
DEFINITION complete cds.
ACCESSION AF529834
VERSION AF529834.1 GI:22297255
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 595)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 595)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..595
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="7"
/cell_type="hybridoma Pl-5"
/transgenic
1..595
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>595
/gene="AID"
1..>595
/gene="AID"
CDS
1..>595
/codon_start=1
/product="AID"
/protein_id="AA095421.1"
/db_xref="GI:22297256"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDRATSFSL
DFGLRNKNGCHVELFLRYISDWLDPPGRYRVTFWTSWSPCYDCARHVADEFLRGNP
NLSLRIFTARLYFCEDRKAEPLRLRLHRAVGQIAIMTFKDYFCWNTEFVENHERTFK
AWEGLEHNSVRLSRQLRILLPLYEVDLDRFRTLGL"
BASE COUNT 127 a 163 c 154 g 151 t
ORIGIN
Alignment Scores:
Pred. No.: 9.57e-149 Length: 595
Score: 149.00 Matches: 197
Percent Similarity: 98.99% Conservative: 0
Best Local Similarity: 98.99% Mismatches: 1
Query Match: 75.25% Indels: 2
DB: 12 Gaps: 0
US-09-966-880A-8 (1-198) x AF529834 (1-595)
Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
Qy 21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAGGGTCGGCTGAGACCTACCTGTGTACGTAGTGAAGAGCGGTGACAGTCTACA 120
Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60

Db 121 TCCCTTTTACGGGACTTTGGTTATCTTCGCAATAAGACGGCTCCACGTGGAAATTCCTC 180
Qy 61 PheLeuArgTyrIleSerAspTyrAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCCTCCGCTACATCTCGGACTGGAGCTAGACCTGGCCGCTGCTACCGGCTCACCTGG 240
Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCCGACATGTGGCCGACTTTCGCGAGG 300
Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCACTCTGAGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGAAG 360
Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCCGAGGGCTCGCGGCTGCACCGCCGGGTGCAATAGGCATCATGACC 420
Qy 141 PheLysAspTyrPheTyrCysTyrAsnThr-PheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAAGATTATTTTACTGCTGGAATAT-TTTTGTAGAAACCATGAAAGAACTTTCAA 479
Qy 160 SalATrPGLuGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
Db 480 AGCCTGGGAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCT 539
Qy 180 uLeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 540 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTT 594

RESULT 24
AF529836
LOCUS
DEFINITION Mus musculus clone 9 transgenic Homo sapiens AID (AID) mRNA, complete cds.
ACCESSION AF529836
VERSION AF529836.1 GI:22297259
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
JOURNAL Somatic hypermutation of the AID transgene in B and non-B cells
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
JOURNAL Direct Submission
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES
source

RESULT 25
AF529853

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

BASE COUNT 127 a 164 c 153 g 152 t
ORIGIN
Alignment Scores:
Pred. No.: 1.48e-145 Length: 596
Score: 146.00 Matches: 176
Percent Similarity: 98.88% Conservative: 0
Best Local Similarity: 98.88% Mismatches: 1
Query Match: 73.74% Indels: 2
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529836 (1-596)

Qy 22 LysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSer 41
Db 64 AAGGTCGGGCTGAGACCTACCTGTCTAGTGAAGAGCGCTGACATGCTACATACC 123
Qy 42 PheSerLeuAspPheGlyTyrLeuArgAsnLys-AsnGlyCysHisValGluLeuLeuPhe 61
Db 124 TTTTCACTGGACTTTGGTTATCTTCGCAATAA-TAACGGCTGCCACCTGCAATTCCTT 182
Qy 61 eLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPh 81
Db 183 CCTCCGCTACATCTCGGACTGGACCTAGACCTGGCCGCTGCTACCGCTTCACCTGGT 242
Qy 81 eThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAs 101
Db 243 CACCTCTCGAGGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTTGGAGGGAA 302
Qy 101 nProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAl 121
Db 303 CCCCAACCTCAGTCTCAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAGC 362
Qy 121 aGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPh 141
Db 363 TGAGCCCGAGGGCTGCGGGGCTGCACCGCGCGGCTGCAATAGATGATCATGACTT 422
Qy 141 eLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAl 161
Db 423 CAAAGATTATTTTACTGCTGGAATACTTTGTAGAAACCATGAAAGAACTTTCAANGC 482
Qy 161 aTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLe 181
Db 483 CTGGGAAGGCTCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGGCGATCCTTTT 542
Qy 181 uProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 543 GCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTT 594

AF529853 597 bp mRNA linear SYN 19-AUG-2002
Cricetulus griseus clone 13 transgenic Homo sapiens AID (AID) mRNA, complete cds.

ACCESSION AF529853

VERSION AF529853.1 GI:22297293

KEYWORDS

SOURCE Chinese hamster.

ORGANISM Cricetulus griseus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;

Cricetulus.

REFERENCE 1 (bases 1 to 597)

AUTHORS Martin, A. and Scharff, M.D.

TITLE Somatic hypermutation of the AID transgene in B cells and non-B

JOURNAL cells

REFERENCE 2 (bases 1 to 597)

AUTHORS Martin, A. and Scharff, M.D.

TITLE Direct Submission

JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of

Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

Db	540	TTTGCCCTGTATGAGCTTGATGACTTACGAGACGCACTTCGTACT	585
FEATURES	Location/Qualifiers		
source	1..597		
	/organism="Cricetulus griseus"		
	/db_xref="taxon:10029"		
	/clone="13"		
	/cell_line="CHO"		
	/transgenic		
source	1..597		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
gene	1..597		
	/gene="AID"		
CDS	1..597		
	/gene="AID"		
	/codon_start=1		
	/product="AID"		
	/protein_id="AAM95440.1"		
	/db_xref="GI:22297294"		
	/translation="MDSLMMRRKFLYQFNVRWAKGRRETYLCYVVVKRDSATSPSL		
	DFGYLRNKGCHVELLFURYISDMDLPGRCYRVTFWTSWSPCYDCARHVADFLRGNP		
	NLSLRIFATFLYFCEDRKAEPLRLHRAGVQIAIMTFKDYFYWNTEFVENHERTFK		
	AWEGHLHNSVRLSRQLRILLPLYEVDLDRDAFTWGR"		
BASE COUNT	129 a 164 c 156 g 148 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1..48e-145	Length:	597
Score:	146.00	Matches:	194
Percent Similarity:	98.98%	Conservative:	0
Best Local Similarity:	98.98%	Mismatches:	1
Query Match:	73.74%	Indels:	2
DB:	12	Gaps:	0
US-09-966-880A-8 (1-198) x AF529853 (1-597)			
QY	1	MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp	20
Db	1	ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAAAATGTCACGCTGG	60
QY	21	AlaLysGlyArgArgGluThrTyrLeuCystYrValValLysArgArgAspSerAlaThr	40
Db	61	GCTAAGGTCGCGTCAGACCTACTCTGCTAGCTAGTAGAGAGCGGTGACAGTGCTACA	120
QY	41	SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu	60
Db	121	TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGCTGCCACGTGGAATTCCTC	180
QY	61	PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp	80
Db	181	TTTCCTCCGCTACATCTCGAGCTGGGACCTAGACCTTGGCCGCTGCTACCGCGTCACCTGG	240
QY	81	PheThrSerTrpSerProCystYrAspCysAlaArgHisValAlaAspPheLeuArgGly	100
Db	241	TTACCTCTCTGAGCCCTCTGCTACGACATGTCCCGACATGTGGCCGACTTTCTGCGAGGG	300
QY	101	AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys	120
Db	301	AACCCCAACCTCAGTCTGAGGATCTTACCAGCGCGCTCTACTTCTGTGAGGACCGCAAG	360
QY	121	AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr	140
Db	361	GCTAGCCCGAGGGCTCGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC	420
QY	141	PheLysAspTyrPheTyrCys-TrpAsnThrPheValGluAsnHisGluArgThrPheLy	160
Db	421	TTCAAGAATATTTTACTTA-CTGGAATACTTTTGTAGAAAACCATGAAGAATTTCAA	479
QY	160	sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe	180
Db	480	AGCCTGGGAGGGCTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCT	539
QY	180	uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThr	195
FEATURES	Location/Qualifiers		
source	1..597		
	/organism="Cricetulus griseus"		
	/db_xref="taxon:10029"		
	/clone="13"		
	/cell_line="CHO"		
	/transgenic		
source	1..597		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
gene	1..597		
	/gene="AID"		
CDS	1..597		
	/gene="AID"		
	/codon_start=1		
	/product="AID"		
	/protein_id="AAM95440.1"		
	/db_xref="GI:22297294"		
	/translation="MDSLMMRRKFLYQFNVRWAKGRRETYLCYVVVKRDSATSPSL		
	DFGYLRNKGCHVELLFURYISDMDLPGRCYRVTFWTSWSPCYDCARHVADFLRGNP		
	NLSLRIFATFLYFCEDRKAEPLRLHRAGVQIAIMTFKDYFYWNTEFVENHERTFK		
	AWEGHLHNSVRLSRQLRILLPLYEVDLDRDAFTWGR"		
BASE COUNT	129 a 164 c 156 g 148 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1..48e-145	Length:	597
Score:	146.00	Matches:	194
Percent Similarity:	98.98%	Conservative:	0
Best Local Similarity:	98.98%	Mismatches:	1
Query Match:	73.74%	Indels:	2
DB:	12	Gaps:	0
US-09-966-880A-8 (1-198) x AF529853 (1-597)			
QY	1	MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp	20
Db	1	ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAAAATGTCACGCTGG	60
QY	21	AlaLysGlyArgArgGluThrTyrLeuCystYrValValLysArgArgAspSerAlaThr	40
Db	61	GCTAAGGTCGCGTCAGACCTACTCTGCTAGCTAGTAGAGAGCGGTGACAGTGCTACA	120
QY	41	SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu	60
Db	121	TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGCTGCCACGTGGAATTCCTC	180
QY	61	PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp	80
Db	181	TTTCCTCCGCTACATCTCGAGCTGGGACCTAGACCTTGGCCGCTGCTACCGCGTCACCTGG	240
QY	81	PheThrSerTrpSerProCystYrAspCysAlaArgHisValAlaAspPheLeuArgGly	100
Db	241	TTACCTCTCTGAGCCCTCTGCTACGACATGTCCCGACATGTGGCCGACTTTCTGCGAGGG	300
QY	101	AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys	120
Db	301	AACCCCAACCTCAGTCTGAGGATCTTACCAGCGCGCTCTACTTCTGTGAGGACCGCAAG	360
QY	121	Al	


```

source
1..597
/transgenic
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/clone="3"
/feature="AID"
1..597
/feature="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95428.1"
/db_xref="GI:22297270"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELFLRYISDMDLDPGRCYRVWTFTSPSPCDCAHVAADFRLGNP
NLSLRITFARLYFCEDRKAEPGLRLHRAGVQIAIMTKDYFCWNTFVENHRTFK
AWEGLHNSVRLSQLRILLPLVEVDLDRADFRTWGR"
BASE COUNT 129 a 164 c 157 g 147 t
ORIGIN

Alignment Scores:
Pred. No.: 3.52e-139 Length: 597
Score: 140.00 Matches: 194
Percent Similarity: 98.98% Conservative: 0
Best Local Similarity: 98.98% Mismatches: 1
Query Match: 70.71% Indels: 2
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529841 (1-597)

QY 1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
DB 1 ATGGACAGCCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
DB 61 GCTAAGGTCGCGGTGAGACCTACTGTGCTACGTAGTGAAGCGCGGTGACAGTGTCTACA 120
QY 41 SerPheSerLeuAspPhePheThrAlaArgLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
DB 121 TCCCTTTTCACTGGACTTCGAGACCTTTCACCAATAAGACGCGTGCACGATGCTGCTC 180
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspPheLeuArgCysTyrArgValThrTrp 80
DB 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGGCTCACCTGG 240
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
DB 241 TTCACCTCTGGAGCCCTGTACGACTGTGCGCGACATGTGGCGGCTTTCTGCGAGGG 300
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
DB 301 AACCCCAACCTCAGTCTCAGGACTTTCACCGCGCGGCTTACTTCTGTGAGACCGCAAG 360
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
DB 361 GCTGAGCCGAGGGGCTGCGCGGCTGACCGCGCGGCGGTGCAATAGCCATCATGACC 420
QY 141 Phe-LysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160
DB 421 AT-CANAGATTATTTTACTGCTGGAATACITTTGTGAAACCAACCTGAAAGAACTTTCAA 479
QY 160 sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
DB 480 AGCCTGGGAAGGCTGCATGAAATTCAGTTTCGTCTCCAGACAGCTTCGCGCATCCT 539
QY 180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThr 195
DB 540 TTTGCCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTCTGCTACT 585

RESULT 29
AF529830 596 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 3 transgenic Homo sapiens AID (AID) mRNA,

```

```

complete cds.
AF529830
AF529830.1 GI:22297247
house mouse.
Mus musculus
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 596)
Martin,A. and Scharff,M.D.
Somatic hypermutation of the AID transgene in B and non-B cells
Unpublished
2 (bases 1 to 596)
Martin,A. and Scharff,M.D.
Direct Submision
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
location/Qualifiers
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="3"
/cell_type="hybridoma PI-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/feature="AID"
1..>596
/feature="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95417.1"
/db_xref="GI:22297248"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELFLRYISDMDLDPGRCYRVWTFTSPSPCDCAHVAADFRLGNP
NLSLRITFARLYFCEDRKAEPGLRLHRAGVQIAIMTKDYFCWNTFVENHRTFK
AWEGLHNSVRLSQLRILLPLVEVDLDRADFRTLGL"
BASE COUNT 126 a 164 c 155 g 151 t
ORIGIN

Alignment Scores:
Pred. No.: 4.06e-138 Length: 596
Score: 139.00 Matches: 197
Percent Similarity: 98.99% Conservative: 0
Best Local Similarity: 98.99% Mismatches: 1
Query Match: 70.20% Indels: 2
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529830 (1-596)

QY 1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
DB 1 ATGGACAGCCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGCTGG 60
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
DB 61 GCTAAGGTCGCGGTGAGACCTACTGTGCTACGTAGTGAAGCGCGGTGACAGTGTCTACA 120
QY 41 SerPheSerLeuAspPhePheThrAlaArgLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
DB 121 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGCGTGCACGATGCTGCTC 180
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspPheLeuArgCysTyrArgValThrTrp 80
DB 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGGCTCACCTGG 240
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
DB 241 TTCACCTCTGGAGCCCTGTACGACTGTGCGCGACATGTGGCGGCTTTCTGCGAGGG 300
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120

```



```

QY      140 rPhelyAspTyrPheTyrCysTirPAsnThrPheValGluAsnHisGluArgThrPhely 160
      |||
      420 CTTCAAGATATATTTTACTGCTGGAATCTTTGTAGAAAACCATGAAGAACTTTCAA 479
      |||
QY      160 salATrPGLuGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgileLe 180
      |||
      480 AGCCTGGGAAGGCTGCATGAAATTCAGTTGCTCTCCACAGACAGCTTCGGCGCATCCT 539
      |||
QY      180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
      |||
      540 TTTGCCCTGTATGAGTGTAGTACTTACGAGACCATTTCTGACTTTGGGACTT 594

RESULT 33
AF529848      597 bp      mRNA      linear      SYN 19-AUG-2002
LOCUS      Cricetulus griseus clone 8 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION      complete cds.
ACCESSION      AF529848
VERSION      AF529848.1 GI:22297283
KEYWORDS      Chinese hamster.
SOURCE      Cricetulus griseus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
      Cricetulus.
REFERENCE      1 (bases 1 to 597)
AUTHORS      Martin,A. and Scharff,M.D.
TITLE      Somatic hypermutation of the AID transgene in B cells and non-B
      cells
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 597)
AUTHORS      Martin,A. and Scharff,M.D.
TITLE      Direct Submission
JOURNAL      Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
      Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES      Location/Qualifiers
      source      1..597
      /organism="Cricetulus griseus"
      /db_xref="taxon:10029"
      /clone="8"
      /cell_line="CHO"
      /transgenic
      1..597
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      1..597
      /gene="AID"
      1..597
      /gene="AID"
      /codon_start=1
      /product="AID"
      /protein_id="AA095435.1"
      /db_xref="GI:22297284"
      /translation="MDSLMMRRFLYQFKNVRWAKGRRETYLCYVVKRRDSATFSFL
      DFGYLRNKGHVELLYRISDWGLDPGRYRVYFTWTSWSPCYDCARHVAADFLRGNP
      NLSRIFARLYFEDRAEPEGLRLRHAGVOIAIMTFKDYFCWNTFVENHETFK
      AWGLHENSVRSLRRLPLLYEVDLRDAFTWGR"
BASE COUNT      127 a 164 c 158 g 148 t
ORIGIN

Alignment Scores:
Pred. No.:      2,66e-124      Length:      597
Score:      126.00      Matches:      194
Percent Similarity:      98.98%      Conservative:      0
Best Local Similarity:      98.98%      Mismatches:      1
Query Match:      63.64%      Indels:      2
DB:      12      Gaps:      0

US-09-966-880A-8 (1-198) x AF529848 (1-597)

QY      1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
      |||

```

```

Db      1 ATGCAGAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATCGCGCTGG 60
QY      21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
      |||
      61 GCTAAGGTCGCGGAGACCTTACCTGCTACGTACGTACGTAGTAGAAGGCGGTGACAGTCTACA 120
      |||
QY      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
      |||
      121 TCCTTTTTCACGTGACATTTGGTTATCTTCCCAATAAGAACGGCTGCCACCTGGAATTCGTC 180
      |||
QY      61 PheLeuArgTyrIleSerAspTrpAsp-LeuAspProGlyArgCysTyrArgValThrTr 80
      |||
      181 TTCTCCCGCTACATCTCGACTGGGG-CCTAGACCTCGCGCTGCTACCGGCTCACCTG 239
      |||
QY      80 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArg 100
      |||
      240 GTTCACCTCCCTGGAGCGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGCGAGG 299
      |||
QY      100 YasnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArg 120
      |||
      300 GAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAA 359
      |||
QY      120 sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
      |||
      360 GGCTGAGCCCGAGGGGCTCGCGGCTGCACCGCGCGGGTGCAAATAGTCATCATGAC 419
      |||
QY      140 rPhelyAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPhely 160
      |||
      420 CTTCAAGATATATTTTACTGCTGGAATCTTTGTAGAAAACCATGAAGAACTTTTCAA 479
      |||
QY      160 salATrPGLuGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgileLe 180
      |||
      480 AGCCTGGGAAGGCTGCATGAAATTCAGTTGCTCTCCACAGACAGCTTCGGCGCATCCT 539
      |||
QY      180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThr 195
      |||
      540 TTTGCCCTGTATGAGTGTAGTACTTACGAGACCATTTCTGACTTTTCGTA 585
      |||

RESULT 34
AF529839      596 bp      mRNA      linear      SYN 19-AUG-2002
LOCUS      Mus musculus clone 12 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION      complete cds.
ACCESSION      AF529839
VERSION      AF529839.1 GI:22297265
KEYWORDS      house mouse.
SOURCE      Mus musculus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
      Martin,A. and Scharff,M.D.
REFERENCE      1 (bases 1 to 596)
AUTHORS      Martin,A. and Scharff,M.D.
TITLE      Somatic hypermutation of the AID transgene in B and non-B cells
      Unpublished
JOURNAL      2 (bases 1 to 596)
REFERENCE      Martin,A. and Scharff,M.D.
AUTHORS      Direct Submission
TITLE      Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
      Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
JOURNAL      Location/Qualifiers
FEATURES      source
      1..596
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /clone="12"
      /cell_type="hybridoma p1-5"
      /transgenic
      1..596
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      1.->596
      /gene="AID"
      1.->596
      /gene="AID"

```

```
/codon_start=1
/product="AID"
/protein_id="AAM95426.1"
/db_xref="GI:22297266"
/translation="MDSLMMNRKFLYQFKNVRKAGRPETYLCYVVKRRDSATSFSL
DFGYLRKNGCHVELLFRLYISDWDLDGRCYRVYTWTSNPSPCYDCARHVADEFIRGNP
NLSLRITFARLYCFEDRKAEPGLRHRAGVQIAIMTFKDYFCWNTFVENHBERTEK
AMEGLHNSVRLSRQLRLLPLYEVDLDRDAFRTLGL"
BASE COUNT      128 a 163 c 155 g 150 t
ORIGIN
Alignment Scores:
Pred. No.:      4.74e-120      Length:      596
Score:          122.00      Matches:      197
Percent Similarity: 98.98%      Conservative: 0
Best Local Similarity: 98.99%      Mismatches: 1
Query Match:      61.62%      Indels:      2
DB:              12      Gaps:      0

US-09-966-880A-8 (1-198) x AF529839 (1-596)
Qy      1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db      1 ATGGACACCCCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAAATGTCGCGTGG 60
Qy      21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db      61 GCTAAGGGTCGCGGTGACACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 120
Qy      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db      121 TCCCTTTTCACGTGGACTTTGGTATATCTTCCCAATAAGAACGCGCTGCCACGTGGAATTGCTC 180
Qy      61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db      181 TTCCTCCGCTACATCTCGGACTGGGACCTACACCTGCGCGCTGCTACCGGTCACTCGG 240
Qy      81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db      241 TTCACCTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGCGAGGG 300
Qy      101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db      301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360
Qy      121 AlaGluPro-GluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
Db      361 CTGAGAC-CGAGGGGCTGCGGGCTGCACCGCGCGGGGTGCNAATAGCCATCATGAC 419
Qy      140 rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160
Db      420 CTTCAAGATTATTTTACTGCTGGAATACTTTTGTAAGAAACCATGAAAGAACTTTCAA 479
Qy      160 salatrPGLuGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
Db      480 AGCCTGGGAAGGGTGCATGAAATTCAGTTTCGTCTCTCCAGACAGCTTCGCGCGCATCCT 539
Qy      180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db      540 TTTGCCCTGTATGAGGTGTGACTTACGAGACGCAATTCGTACTTTGGGACTT 594

RESULT 35
AF529850
LOCUS
DEFINITION
Cricetulus griseus clone 10 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
ACCESSION
AF529850
VERSION
AF529850.1
KEYWORDS
GI:22297287
SOURCE
Chinese hamster.
ORGANISM
Cricetulus griseus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
```

```
QY 140 rPhelysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160
|||||
Db 420 CTTCAAGATTATTTTACTGCTGGATACTTTTGTAGAAAACCATGAAGAACTTTCAA 479

QY 160 salaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgileLe 180
|||||
Db 480 AGCCTGGGAAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCT 539

QY 180 uLeuProLeuTyrGluValAspLeuArgAspAlaPheArgThr 195
|||||
Db 540 TTTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACT 585

RESULT 36
AF529818
LOCUS Homo sapiens clone Ramos 4 truncated AID (AID) mRNA, complete cds. PRI 19-AUG-2002
DEFINITION AF529818
ACCESSION AF529818
VERSION AF529818.1 GI:22297223
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 4"
1..596
/gene="AID"
1..570
/gene="AID"
/feature="AID"
/feature="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="truncated AID"
/protein_id="AAM95405.1"
/db_xref="GI:22297224"
/translation="MDSLMMRRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFRLYISDWLDPGRCYRVYTWFTSWSPCYDCARHVADFLRGNP
NLSRIETARLYFCEDHKAPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHETFK
AWEGIHENSRLSLRLRLPLVEVDL"
BASE COUNT 128 a 163 c 154 g 151 t
ORIGIN

Alignment Scores:
Pred. No.: 8.45e-116 Length: 596
Score: 118.00 Matches: 188
Percent Similarity: 98.95% Conservative: 0
Best Local Similarity: 98.95% Mismatches: 1
Query Match: 59.60% Indels: 2
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AF529818 (1-596)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
|||||
Db 1 ATGGACAGCCCTTGTATGAACCGGAGGAGTTCTTTACCAATTCAAAAATGTCGCTGG 60

QY 21 AlaLysGlyArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
|||||
Db 61 GCTAAGGTCGCGCTGAGACCTACTCTGCTACGTAGTAGAAGAGCGCTGACAGTGCTACA 120
```

```
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
|||||
Db 121 TCCTTTTCTACTGGACTTTGGTTATCTTCCCAATAAGAACGGCTGCCACCTGGAATTC 180

QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
|||||
Db 181 TTCCCTCCGCTACATCTCGACTGGGACCTAGACCCTGCCCGCTGCTACCGCTACCTGG 240

QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
|||||
Db 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCCGACATGTGCCGCACTTTCTGCGAGGG 300

QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLy 120
|||||
Db 301 AACCCCAACCTCAGTCTCAGGATCTTACCCGCGCCCTCTACTCTCTGTGAGACCA-CAA 359

QY 120 salaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
|||||
Db 360 GGCTGAGCCGAGGGCTGCGCGGCTGCACCGCGCGGTGCAATAGCCATCATGAC 419

QY 140 rPhelysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160
|||||
Db 420 CTTCAAGATTATTTTACTGCTGGATACTTTTGTAGAAAACCATGAAGAACTTTCAA 479

QY 160 salaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgileLe 180
|||||
Db 480 AGCCTGGGAAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCT 539

QY 180 uLeuProLeuTyrGluValAspLeuArgAspAlaPheArgThr 195
|||||
Db 540 TTTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACT 585

RESULT 37
AF529819
LOCUS Homo sapiens clone Ramos 5 AID (AID) mRNA, partial cds. PRI 19-AUG-2002
DEFINITION AF529819
ACCESSION AF529819
VERSION AF529819.1 GI:22297225
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 5"
1..596
/gene="AID"
1..596
/gene="AID"
/feature="AID"
/feature="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95406.1"
/db_xref="GI:22297226"
/translation="MDSLMMRRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFRLYISDWLDPGRCYRVYTWFTSWSPCYDCARHVADFLRGNP
NLSRIETARLYFCEDHKAPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHETFK
AWEGIHENSRLSLRLRLPLVEVDL"
BASE COUNT 128 a 164 c 154 g 150 t
ORIGIN
```


Db	241	TTCACCTCCCTGGAGCCCTGCTACGACTGTGCCCGACATGTGCCGACGTTTCCTGCCAGGG	300
Qy	101	AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGlu-AspArgL	120
Db	301	AACCCCAACCTCAGTCGAGATCTTCACCGCGGCGCTCTACTTCTGTG-GGACCGCAA	359
Qy	120	sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh	140
Db	360	GGCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGGTGCAAAATACCATCATGAC	419
Qy	140	rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy	160
Db	420	CTTCAAGAGATTATTTTACTCTCGAATACTTTTGTAGAAACCATCAAGAACTTTTCAA	479
Qy	160	sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGluLeuArgGluLeu	180
Db	480	AGCGTGGAGAGGCTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTGGCGCATCTT	519
Qy	180	uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThr	195
Db	540	TTTGCCCTGTATGAGGTTGATGACTTACGAGACGATTCGTACT	585
RESULT 40			
AF529840			
LOCUS		577 bp	mRNA linear SYN 19-AUG-2002
DEFINITION		Mus musculus clone 13 transgenic Homo sapiens AID (AID) mRNA,	
ACCESSION		complete cds.	
VERSION		AF529840	
KEYWORDS		AF529840.1 GI:2297267	
SOURCE		house mouse.	
ORGANISM		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;	
TITLE		Martin,A. and Scharff,M.D.	
JOURNAL		Somatic hypermutation of the AID transgene in B and non-B cells	
REFERENCE		Unpublished	
AUTHORS		2 (bases 1 to 577)	
TITLE		Martin,A. and Scharff,M.D.	
JOURNAL		Direct Submission	
FEATURES		Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of	
source		Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA	
		Location/Qualifiers	
		1..577	
		/organism="Mus musculus"	
		/db_xref="taxon:10090"	
		/clone="13"	
		/cell_type="hybridoma pl-5"	
		/transgenic	
		1..577	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		<1..>577	
		/gene="AID"	
		<1..>577	
		/gene="AID"	
		/codon_start=3	
		/product="AID"	
		/protein_id="AA095427.1"	
		/db_xref="GI:2297268"	
		/translation="RKFLYQPKNVWAKGRRETYLCYVVKRDSATSFSLDPGYLRN	
		KNGCHEVLLFLRYISDWDDPGRCYRVFTWSKPCYDCARHVAFLRGNPNLSRLIR	
		TARLYFCEDORKAESGLRHAGVOIAITMTFDYFCWNTFVENHRTFKAWEGHLE	
		NSVRLSRQLRRILLPLYEVDLURDAFTLGL"	
BASE COUNT	122 a	159 c	150 g 146 t
ORIGIN			
Alignment Scores:			
Pred. No.:	1.27e-112	Length:	577
Score:	115.00	Matches:	190
Percent Similarity:	98.96%	Conservative:	0
Best Local Similarity:	98.96%	Mismatches:	1

```
Query Match: 58.08% Indels: 2
DB: 12 Gaps: 0
US-09-966-880A-8 (1-198) x AF529840 (1-577)

QY 8 ArqArgLysPheLeuTyrglnPheLysAsnValArgTrpAlaLysGlyArgArgGluThr 27
D 3 CGAGAGAGTTCCTTTACCAATTCAAAAATGCCCTGGCTGAAGGTCGGCGTGAGACC 62
QY 28 TyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPheSerLeuAspPheGly 47
D 63 TACCTGTCTAGTAGTAGAGAGCGGTGACAGTGTATCATCTTTTCACTGAGCATTTGGT 122
QY 48 TyrLeuArgAsnLysAsnGlyCysHisValGluLeuPheLeuArgTyrIleSerAsp 67
D 123 TATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCCTTCTCCGCTACATCTCGGAC 182
QY 68 TrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCys 87
D 183 TGGGACCTAGACCCCTGGCGCGCTGCTACCGCGTCACCTGGTTCCACTCTGGAGCCCTGC 242
QY 88 TyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArg 107
D 243 TACGACTGTGCCGACATGTGCCGACTTCTGCGAGGGAACCCCAACCTCAGTCTGAGG 302
QY 108 IlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluPro-GluGlyLeuAr 127
D 303 ATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGTC-CGAGGGGCTGCG 361
QY 127 gArGluHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyrCy 147
D 362 GCGGCTGACCGCGCGGGGTGCAATAGCCATCATGACCTTCAAGAGATTATTTTACTG 421
QY 147 sTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGlyLeuHisGl 167
D 422 CTGGATACTTTGTAGAAACCATGAAGAATTTCAAGGCTGGGAGGCTGCATGA 481
QY 167 uAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLeuLeuProLeuTyrGluValAs 187
D 482 AAATTCACTGCTCTCCAGACAGCTTCGGCGCATCTTTTGCCCTGTATGAGTTGA 541
QY 187 pAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
D 542 TGACTTACGAGACGCATTTCTGACTTTGGGACTT 575

RESULT 41
AF529816
LOCUS Homo sapiens clone Ramos 2 AID (AID) mRNA, linear PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 2 AID (AID) mRNA, partial cds.
ACCESSION AF529816
VERSION AF529816.1 GI:22297219
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLES Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLES Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Bronx, NY 10461, USA
FEATURES
source location/Qualifiers
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 2"
1..>596
/gene="AID"
```

```
CDS
1..>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95403.1"
/db_xref="GI:22297220"
/translation="MDSLMMNRKFLYQKNVRWAKGRRETYLCVYVRRDSATSFSL
DFGLRNKNGCHVELLFURYISDMDLPGRCYRVTFWTSRSPCYDCARHVDLFRGNP
NLSRIETARLYFCEDRKAEPGLRLRHAGVQIAITMFKDYFYCWNTFVENHRTFK
AWEGHENSVRLSROLRLILLPLYEVDLROAFRTLGL"
BASE COUNT 127 a 165 c 155 g 149 t
ORIGIN
Alignment Scores:
Pred. No.: 151e-111 Length: 596
Score: 114.00 Matches: 197
Percent Similarity: 98.99% Conservative: 0
Best Local Similarity: 98.99% Mismatches: 1
Query Match: 57.58% Indels: 2
DB: 9 Gaps: 0
US-09-966-880A-8 (1-198) x AF529816 (1-596)
QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
D 1 ATGGACAGCCTCTTATGAAACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGCTGG 60
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
D 61 GCTAAGGGTCGGCGGTGAGACCTACCTGTCTACGTAGTGAAGAGCGGTGACAGTCTACA 120
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
D 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAAGCGCTGCCACGTGGAATGGTC 180
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
D 181 TTCCCTCCGCTACATCTCGACTGGGACCTAGACCTGGCGCTGTACCGCTCACCTGG 240
QY 81 PheThrSerTrp-SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGl 100
D 241 TTCACCTCCCG-GAGCCCTCTGTAGACTGTGCCGACATGTGCCGACTTTTCTCGGAGG 299
QY 100 YAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgly 120
D 300 GAACCCCAACCTCAGTCTGAGGATCTTACCCTGGCGCTCTACTTCTGTGAGGACCGCAA 359
QY 120 sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140
D 360 GGCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGGTGCAAAATAGCCATCATGAC 419
QY 140 rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160
D 420 CTTCAAGAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCATGAAGAACCTTTCAA 479
QY 160 sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180
D 480 AGCCTGGGAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGCGCGCATCT 539
QY 180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
D 540 TTTGCCCTGTATGAGTTGATGACTTACGAGACGCAITTCGTACTTTTGGGACTT 594
RESULT 42
AF529827
LOCUS Homo sapiens clone Ramos 13 AID (AID) mRNA, linear PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 13 AID (AID) mRNA, partial cds.
ACCESSION AF529827
VERSION AF529827.1 GI:22297241
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES
source
1..596
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 13"
1..>596
/gene="AID"
1..>596
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95414.1"
/db_xref="GI:22297242"
/translation="MDSLMMNRKFLYOFKNVRWAKGRRTYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFLRYISDNDLDPGRCYRVTWFTSWSPCYDCARHVDPLRGNP
NLSLRIFTARLYFCEDRKAEPGLRLRHAGVQVTAIMTFKDYFCWNTFVENHERTFK
AWEGHENSVRSLRQLRLLPLYEVDLDRADFRTWGR"

BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN

Alignment Scores:
Pred. No.: 3 1e-106 Length: 596
Score: 109.00 Matches: 197
Percent Similarity: 98.99% Conservative: 0
Best Local Similarity: 98.99% Mismatches: 1
Query Match: 55.05% Indels: 2
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AF529827 (1-596)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
DB 1 ATGGACAGCCTCTTATGATGAACGGAGGAGTCTCTTACCAATTCAAAATGTCGCTGG 60
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
DB 61 GCTAAGGTCGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTACA 120
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
DB 121 TCCTTTTCACGGACTTTGGTTATCTCCGAATAAGACGCTGCCACGTGAATGCTC 180
QY 61 PheLeuArgTyrTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
DB 181 TTCCTCCGCTACATCTCGAGCTGGGACCTAGACCTGCGCGCTGCTACCGGTCACTGG 240
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
DB 241 TTCACCTCCTGAGCCCTGCTAGACATCTGTCGCGACATGTGGCGGACTTTCTCGGAGGG 300
QY 101 AsnProAsnLeuSerLeuArgIlePheThr-AlaArgLeuTyrPheCysGluAspArgLy 120
DB 301 AACCCCAACCTCAGTCTGAGGATCTTCAT-CGCGCGCTCTACTCTGTGTGAGGACCGCAA 359
QY 120 sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMeth 140
DB 360 GCCTGAGCCGAGGGGCTGCGCGCGCTGCACCGCGCGGGTGCAAATAGCCATCATGAC 419
QY 140 rPheLysAspTyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPheLy 160
|||||

DB 420 CTTCAAAGATTATTTTACTGCTGCAATCTTTCTAGAAAAACCATGAAGAATCTTCAA 479
QY 160 sLatrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
DB 480 AGCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGATCCT 539
QY 180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
DB 540 TTTGCCCTCTGATGAGTTGATGACTTACGAGACGCATTTCTGACTTTTGGGACTT 594

RESULT 43
AF529854
LOCUS
DEFINITION
Cricetulus griseus clone 14 transgenic Homo sapiens AID (AID) mRNA,
complete cds.

ACCESSION AF529854
VERSION AF529854.1 GI:22297295
KEYWORDS
SOURCE Chinese hamster.
ORGANISM Cricetulus griseus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetidae;
Cricetulus.

REFERENCE 1 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Martin, A. and Scharff, M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES
source
1..597
Location/Qualifiers
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/clone="14"
/cell_line="CHO"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95441.1"
/db_xref="GI:22297296"
/translation="MDSLMMNRKFLYOFKNVRWAKGRRTYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFLRYISDNDLDPGRCYRVTWFTSWSPCYDCARHVDPLRGNP
NLSLRIFTARLYFCEDRKAEPGLRLRHAGVQVTAIMTFKDYFCWNTFVENHERTFK
AWEGHENSVRSLRQLRLLPLYEVDLDRADFRTWGR"

source

gene

CDS

BASE COUNT 128 a 163 c 158 g 148 t
ORIGIN
Alignment Scores:
Pred. No.: 3 59e-105 Length: 597
Score: 108.00 Matches: 194
Percent Similarity: 98.98% Conservative: 0
Best Local Similarity: 98.98% Mismatches: 1
Query Match: 54.55% Indels: 2
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529854 (1-597)
QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
DB 1 ATGGACAGCCTCTTATGATGAACGGAGGAGTCTCTTACCAATTCAAAATGTCGCTGG 60
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40

Db	61	GCTAAGGTCGGCGTAGACCTACCTGTCTACGTAGTGAAGAGCGGTGACAGTGTCTACA	120
QY	41	SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu	60
Db	121	TCCTTTTCACTGGACTTGGTTATCTTCGCATAGAAGCGGTGCCACGTGAATGCTC	180
QY	61	PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp	80
Db	181	TTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGGCGCGTGTACCGCTCACCTGG	240
QY	81	PheThrSerTrpSer-ProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgL	100
Db	241	TTTACCTCTCGAGCCCTGG-TAGACTGTGCCGACATGTGGCGGACTTTCTGCGAGG	299
QY	100	YAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgly	120
Db	300	GAACCCCAACCTCAGTCTGAGGATCTTACCAGCGCGCTCTACTTGTGTGAGGACCGCA	359
QY	120	sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetH	140
Db	360	GGCTGAGCCGAGGGCTGCGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGAC	419
QY	140	rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheL	160
Db	420	CTTCAAGATATTTTACTGCTGGAATACTTTGTAGAAACCATGAAGAAGCTTTTCAA	479
QY	160	sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleL	180
Db	480	AGCCTGGGAAGGCTGCATGAANATTCAGTTCGTCTCCACAGACGCTTCGGCGCATCT	539
QY	180	uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThr	195
Db	540	TTTGGCCCTGTATGAGTTGATGACTTACGAGACGCAATTCGTACT	585
RESULT 44			
AF529817			
LOCUS	AF529817	596 bp	mRNA linear PRI 19-AUG-2002
DEFINITION	Homo sapiens clone Ramos 3 AID (AID) mRNA, partial cds.		
ACCESSION	AF529817		
VERSION	AF529817.1	GI:22297221	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 596)		
JOURNAL	Martin, A. and Scharff, M.D.		
FEATURES	Somatic hypermutation of the AID transgene in B cells and non-B		
source	cells		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 596)		
TITLE	Martin, A. and Scharff, M.D.		
JOURNAL	Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of		
FEATURES	Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA		
source	Location/Qualifiers		
gene	1..596		
CDS	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="Ramos 3"		
	1..>596		
	/gene="AID"		
	1..>596		
	/gene="AID"		
	/note="Integrated into Burkitt's lymphoma cell line Ramos"		
	/codon_start=1		
	/product="AID"		
	/protein_id="AAM95404.1"		
	/db_xref="GI:22297222"		
	/translation="MDSLIMRRKFLYOFKNVRWAKRGRTYVLCYVVKRRDSATSFSL		
	DFGYLRNKGCHVELFLRYISDWLDPGRCYRVTFWTSWSPCYDCARHVDILRGNP		
	NLSLRFITARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDDFYCWNTFVNHRTFK		

BASE COUNT	128 a	164 c	156 g	148 t
ORIGIN	AWEGUHENSVRLSRQLRRILLPLYEVDDLLRDAFRTLGL "			
Alignment Scores:				
Pred. No.:	2.03e-92	Length:	596	
Score:	96.00	Matches:	196	
Percent Similarity:	98.00%	Conservative:	0	
Best Local Similarity:	98.00%	Mismatches:	2	
Query Match:	48.48%	Indels:	4	
DB:	9	Gaps:	0	
US-09-966-880A-8 (1-198) x AF529817 (1-596)				
QY	1	MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp	20	
Db	1	ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCTGG	60	
QY	21	AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr	40	
Db	61	GCTAAGGGTCGGCGTGAGACCTACCTGTCTACGTAGTGAAGAGGCGTGACAGTGTCTACA	120	
QY	41	SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu	60	
Db	121	TCCTTTTCACTGGAGCTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAATTGCTC	180	
QY	61	PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp	80	
Db	181	TTCTCTCGGTACATCTCGGACTGGGACCTAGACCTTGGCGCTGTACCGCTCACCTGG	240	
QY	81	PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPhe-LeuArgG1	100	
Db	241	TTCACTCTCTGGAGCCCTGTCTACGACTGTGCCGACATGTGCCGACAT-TCTGCGAGG	299	
QY	100	YasnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLy	120	
Db	300	GAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAA	359	
QY	120	sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh	140	
Db	360	GGCTGAGCCGAGGGCTGCGGGCGCTGCACCGCGCGGGTGCAATAGCCATCATGAC	419	
QY	140	rPheLysAspTyr-PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheL	160	
Db	420	CTTCAAGATGA-TTTTACTGCTGGAATACTTTGTAGAAAAACCATGAAGAAGCTTTCA	478	
QY	160	YsAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleL	180	
Db	479	AGCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCC	538	
QY	180	euLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu	198	
Db	539	TTTTGCCCTGTATGAGTTGATGACTTACGAGACCGCAATTCGTACTTTGGGACTT	594	
RESULT 45				
AB040430				
LOCUS	Homo sapiens AID gene for activation-induced cytidine deaminase,			
DEFINITION	complete cds.			
ACCESSION	AB040430			
VERSION	AB040430.1 GI:9988407			
KEYWORDS	AID; activation-induced cytidine deaminase.			
SOURCE	Homo sapiens DNA.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
TITLE	1 (sites)			
JOURNAL	Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.			
MEDLINE	Isolation, tissue distribution, and chromosomal localization of the			
REFERENCE	human activation-induced cytidine deaminase (AID) gene			
	Genomics 68 (1), 85-88 (2000)			
	20408890			
	2 (sites)			

AUTHORS Revy,P., Muto,T., Levy,Y., Geissmann,F., Plebani,A., Sanal,O., Catalan,N., Forveille,M., Dufourcq-Lagelouse,R., Gennery,A., Tezcan,I., Ersoy,F., Kayserili,H., Ugazio,A.G., Brousse,N., Muramatsu,M., Notarangelo,L.D., Kinoshita,K., Honjo,T., Fischer,A. and Durandy,A.

TITLE Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)

JOURNAL Cell 102 (5), 565-575 (2000)

MEDLINE 20460541

REFERENCE 3 (bases 1 to 11204)

AUTHORS Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.

TITLE Direct Submission

JOURNAL Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan (E-mail:honjoemfour.med.kyoto-u.ac.jp, Tel:81-75-753-4371(ex.4371), Fax:81-75-753-4388)

FEATURES

Location/Qualifiers

Source 1..11204

/organism="Homo sapiens"

/db_xref="taxon:9606"

Join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)

/gene="AID"

Join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)

/gene="AID"

/codon_start=1

/product="activation-induced cytidine deaminase"

/protein_id="BAB12720.1"

/db_xref="GI:9988408"

/translation="MDSLMMNRKFLYQKNRWAKGRREYLCVVKRRDSATSFSLDFGLRNKNGCVHVELFLRYISDMDLPGRCYRVTFWTSWPCDCARHVAADFLRGNP NLSLRIFARLYFCEDRKAEPGLRLRHARGVQTAIMTFKDFYCWNTFVENHRTFK AWEGLHENSVRSLRKLPLLEVDLDRDAFRFLGL"

BASE COUNT 3305 a 2273 c 2373 g 3253 t

ORIGIN

Alignment Scores:

Pred. No.: 6,1e-85 Length: 11204

Score: 90.00 Matches: 90

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 45.45% Indels: 0

DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AB040430 (1-11204)

QY 53 AsnGlyCysHisValGluLeuPheLeuArgTyrIleSerAspTrpAspLeuAspPro 72

DB 7807 AACGGCTGCCAGTGGGAATTCCTCTCCGCTACATCTCGGACTGGGACCTAGACCT 7866

QY 73 GlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArg 92

DB 7867 GCCCGCTGCTACCGCGTACCCTGGTTACCTCTCGAGCCCTGCTACGACTGTGCCCGA 7926

QY 93 HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArg 112

DB 7927 CATGTGGCCGACTTCTCGAGGAGAACCCCAACCTCAGTCTGAGATCTTCACCGCGGC 7986

QY 113 LeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAla 132

DB 7987 CTCACCTCTGTGAGGACCGCAAGCTGAGCCGAGGGCTGCGGCGCTGACCGGCC 8046

QY 133 GlyValGlnIleAlaIleMetThrPheLys 142

DB 8047 GGGGTGCAATAGCCATCATGACCTTCAAA 8076

RESULT 46

AC092184

LOCUS AC092184 Homo sapiens 12 BAC RP11-438L7 (Roswell Park Cancer Institute Human BAC library) complete sequence. PRI 12-JUN-2002

DEFINITION AC092184.7 GI:21206067

ACCESSION AC092184.7

VERSION AC092184.7

KEYWORDS HTG.

SOURCE ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 71132)

REFERENCE

AUTHORS

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 71132)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavey,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Delaney,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delanoy,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Haviak,P., Haves,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Toshikhes,I., Jackson,L.E., Jacobson,B., Jie,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Louised,H., Lozado,R.J., Lu,X., Luciet,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mel,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S., Oguh,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shostari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watling,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kuchelapati,R., Weinstein,G. and Gibbs,R.

2 (bases 1 to 71132)

Unpublished

Worley,K.C.

Direct Submission

Submitted (25-JUN-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 71132)

Unpublished

Worley,K.C.

Direct Submission

Submitted (18-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 71132)

Unpublished

Worley,K.C.

Direct Submission

Submitted (25-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 71132)

Unpublished

Worley,K.C.

Direct Submission

Submitted (12-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On May 25, 2002 this sequence version replaced gi:20901754.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
 entire insert of this clone. Overlapping regions of clones are only
 sequenced and submitted once, so the sequence for the remainder of
 the insert may be found in the record for the adjacent clones.
 Overlapping clones are noted at the beginning and end of the
 Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches
 of a local database that includes entries from dbSTS, GDB, and
 local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
 unpublished.) for Human and Mouse sequences.

Genes and region of sequence similarity are identified by BLAST
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
 EST and cDNA sequences. Genes demonstrate at least two exons
 flanked by consensus splice sites that maintained sequence
 continuity across the splice junctions. Sequences that are not
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
 standard of double strand coverage with a minimum of 2 clones and 2
 reads with no ambiguities or 2 chemistries with a minimum of 2
 clones and 3 reads with no ambiguities. If the sequence quality for
 a region does not meet this standard, it will be indicated in the
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
 standards - estimated error rate less than 1 per 10,000 bases.
 Reports of lowest quality individual bases and measures of base
 quality are listed below. Description of the metrics can be found
 at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES	Location/Qualifiers
Source	1..71132
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosomes="12"
misc_feature	/clone="RP11-438L7"
	complement(1..1448)
	/note="overlaps bases 1..1448 of clone AC092490"
	/function="clone overlap"
STS	187..408
	/standard_name="57233"
STS	439..560
	/standard_name="92005"
repeat_region	774..881
	/rpt_family="MIR"
repeat_region	903..1190
	/rpt_family="AluSq"
repeat_region	1191..1213
	/rpt_family="AT_rich"
STS	1591..1807
	/standard_name="6612"
STS	1744..1819
	/standard_name="8198"
repeat_region	1966..2264
	/rpt_family="AluSq"
repeat_region	3296..3328
	/rpt_family="(TTTC)n"
repeat_region	4097..4249
	/rpt_family="AluSq"
repeat_region	4360..4654
	/rpt_family="AluY"
repeat_region	complement(5140..5262)
	/rpt_family="FLAM_C"
repeat_region	5275..5304

repeat_region	/rpt_family="AT_rich"	3.01e-84	Length:	71132
5305..5563	/rpt_family="AluJo"	90.00	Matches:	90
5564..5599	/rpt_family="AT_rich"	100.00%	Conservative:	0
5644..5733	/rpt_family="MSTD"	100.00%	Mismatches:	0
5734..6012	/rpt_family="AluSq"	45.45%	Indels:	0
6020..6336	/rpt_family="AluX"	9	Gaps:	0
6337..6629	/rpt_family="MSTD"			
6630..6650	/rpt_family="(CAA)n"			
complement(6711..6817)	/rpt_family="MIR"			
complement(7033..7327)	/rpt_family="AluX"			
7607..7745	/rpt_family="FLAM_A"			
7980..8047	/rpt_family="MIR"			
complement(8050..8549)	/rpt_family="LTR47A"			
8567..8698	/rpt_family="MIR"			
complement(8815..9101)	/rpt_family="AluSq"			
complement(11227..12175)	/rpt_family="LTR5"			
STS	12625..12900			
	/standard_name="87432"			
repeat_region	complement(13547..13664)			
repeat_region	complement(13665..13979)			
repeat_region	complement(13980..14193)			
repeat_region	14622..14924			
repeat_region	/rpt_family="AluY"			
repeat_region	complement(15813..17136)			
repeat_region	complement(17272..17395)			
repeat_region	17426..17575			
repeat_region	/rpt_family="MIR"			
repeat_region	19021..19314			

Alignment Scores:
 Pred. No.: 3.01e-84 Length: 71132
 Score: 90.00 Matches: 90
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 45.45% Indels: 0
 DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AC092184 (1-71132)

QY	53	AsnGlyCysHisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspPro	72
Db	42351	AAACGGCTGCACGTGGAAATTGCTCTCCGCTACATCTCGACTGGGACCTAGACCCT	42410
QY	73	GlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArg	92
Db	42411	GGCCGCTGCTACCCGCTCACCTGGTTCACCTCTGGAGCCCTGTGACACGTGCCCCGA	42470
QY	93	HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArg	112
Db	42471	CATGTGGCGGACTTTCTCGAGGGAGAACCCACCTCAGTCTGAGGATCTTCACCGCGCC	42530
QY	113	LeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgLeuHisArgAla	132

Db 42531 CTCTACTTCTGTGAGACCGCAAGGCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCC 42590

QY 133 GlyValGlnIleAlaIleMetThrPhelYs 142

Db 42591 GGGGTGCAATAGCCATCATGACCTTCACA 42620

RESULT 47

AF132979

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-966-880A-8 (1-198) x AF132979 (1-2440)

QY 54 GlyCysHisValGluLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGly 73

Db 252 GGCTGCCAGTGAATTTGTTCTACGCTACATCTCAGACTGGGACCTGGACCGGCGC 311

QY 74 ArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHis 93

Db 312 CGGTGTTACCGGCTCACCTGGTTACCTCCCTGAGCCGCTGCTATGACTGTGCCGCGCAC 371

QY 94 ValAla 95

Db 372 GTGGCT 377

RESULT 48

AC094826

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-966-880A-8 (1-198) x AF132979 (1-2440)

QY 54 GlyCysHisValGluLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGly 73

Db 252 GGCTGCCAGTGAATTTGTTCTACGCTACATCTCAGACTGGGACCTGGACCGGCGC 311

QY 74 ArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHis 93

Db 312 CGGTGTTACCGGCTCACCTGGTTACCTCCCTGAGCCGCTGCTATGACTGTGCCGCGCAC 371

QY 94 ValAla 95

Db 372 GTGGCT 377

AC094826 143044 bp DNA linear HTG 20-DEC-2001
Rattus norvegicus clone CH230-2013, *** SEQUENCING IN PROGRESS ***,
61 unordered pieces.

AC094826

HTG: HTGS-PHASE1

Rattus norvegicus

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 143044)

Murphy D.M., Adams C., Adio-Oduola B., Ali-Osman F.R., Allen C.,

Bentley S.L., Amaralunga H.C., Are J.R., Banks T., Barbieri J.,

Bentley J., Blum K., Blankenburg K., Bonnin D., Bourk J.,

Bowie S., Brive M., Brown E., Brown M., Bryant N.P., Buhay C.,

Burch P., Burkett C., Burrell K.L., Byrd N.C., Carron T.F.,

Carter M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R.,

Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C.,

Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C.,

Davy-Carroll L., Dederich D.A., Delaney K.R., Delgado O.,

Denn A.L., Ding Y., Dinh H.H., Douthwaite K.J., Draper H.,

Dugan-Rocha S., Durbin K.J., Eamhart C., Edgar D., Edwards C.C.,

Elhaj C., Escotto M., Falls T., Ferraguto D., Flagg N., Ford J.,

Foster P., Frantz P., Gabisi A., Gao J., Garcia A., Garner T.,

Garza N., Gill R., Gorrell J.H., Guevara W., Gunaratne P., Hule S.,

Hamilton K., Harris C., Harris K., Hart M., Havlak P., Hawes A.,

Hernandez J., Hernandez O., Hodgson A., Hoques M., Holloway C.,

Hollins B., Homs F., Howard S., Huber J., Hulyk S., Humo J.,

Jackson L.E., Jacobson B., Jia Y., Johnson R., Jollivet S.,

Joudah S., Karlsson E., Kelly S., Khan U., King L., Korvah J.,

Kovar C., Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C.,

Lewis L., Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W.,

Louise H., Lozano R.J., Lu X., Lucier A., Lucier R., Luna R.,

Ma J., Maheshwari M., Mapa P., Martin R., Martindale A.,

Martinez E., Massey E., Mawhney E., McLeod M.P., Meador M.,

Mei G., Metzger M., Miner G., Miner Z., Mitchell T., Mohabbat K.,

Morgan M., Morris S., Moser M., Neal D., Newton J., Newton N.,

Nguyen A., Nguyen N., Nguyen N., Nickerson E., Nwokwu S.,

Ogihara M., Okunou G., Oragunye N., Oviedo R., Pace A., Payton B.,

Peery J., Perez L., Peters L., Pickens R., Primus E., Pu L.L.,

Quiles M., Ren Y., Rives M., Rojas A., Rojubokan I., Rolfe M.,

Ruiz S., Savery G., Scherer S., Scott G., Shen H., Shiohara N.,

Sisson I., Sodergren E., Sonaike T., Sparks A., Stanley H.,

Stone H., Sutton A., Svatek A., Tabor P., Tamerisa A., Tamerisa K.,

Tang H., Tansey J., Taylor C., Taylor T., Telford B., Thomas N.,

Thomas S., Usmani K., Vasquez L., Vera V., Villalon D., Vinson R.,

Wall R., Wang S., Ward-Moore S., Warren R., Washington C.,

Watlington S., Williams G., Williamson A., Wleczek R., Woodson S.,

Worley K., Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Nelson D.,

Weinstock G. and Gibbs R.

Direct Submission

Unpublished

2 (bases 1 to 143044)

Worley K.C.

Direct Submission

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:15624662.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GBMC

Center clone name: CH230-2013

----- Summary Statistics

Assembly program: Phrap: version 0.990329 First call to

findPhrapList

Consensus quality: 110311 bases at least Q40
 Consensus quality: 118777 bases at least Q30
 Consensus quality: 125541 bases at least Q20
 Estimated insert size: 100318; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-1p estimation
 Quality coverage: 1.7x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 61 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 8003: contig of 8003 bp in length
 * 8004 8103: gap of unknown length
 * 8104 12049: contig of 3946 bp in length
 * 12050 12149: gap of unknown length
 * 12150 16674: contig of 4525 bp in length
 * 16675 16774: gap of unknown length
 * 16775 22570: contig of 5796 bp in length
 * 22571 22670: gap of unknown length
 * 22671 27595: contig of 4925 bp in length
 * 27596 27695: gap of unknown length
 * 27696 31790: contig of 4095 bp in length
 * 31791 31890: gap of unknown length
 * 31891 35625: contig of 3735 bp in length
 * 35626 35725: gap of unknown length
 * 35726 38511: contig of 2786 bp in length
 * 38512 38611: gap of unknown length
 * 38612 41020: contig of 2409 bp in length
 * 41021 41120: gap of unknown length
 * 41121 44145: contig of 3025 bp in length
 * 44146 44245: gap of unknown length
 * 44246 46975: contig of 2730 bp in length
 * 46976 47075: gap of unknown length
 * 47076 50825: contig of 3750 bp in length
 * 50826 50925: gap of unknown length
 * 50926 55401: contig of 4476 bp in length
 * 55402 55501: gap of unknown length
 * 55502 58143: contig of 2642 bp in length
 * 58144 58243: gap of unknown length
 * 58244 60960: contig of 2617 bp in length
 * 60961 60960: gap of unknown length
 * 60961 63226: contig of 2266 bp in length
 * 63227 63326: gap of unknown length
 * 63327 66558: contig of 3232 bp in length
 * 66559 69476: gap of unknown length
 * 69477 69576: gap of unknown length
 * 69577 72308: contig of 2732 bp in length
 * 72309 72408: gap of unknown length
 * 72409 74936: contig of 2528 bp in length
 * 74937 75036: gap of unknown length
 * 75037 77300: contig of 2264 bp in length
 * 77301 77400: gap of unknown length
 * 77401 80440: contig of 3040 bp in length
 * 80441 80540: gap of unknown length
 * 80541 83286: contig of 2746 bp in length
 * 83287 83386: gap of unknown length
 * 83387 85219: contig of 1833 bp in length
 * 85220 85319: gap of unknown length
 * 85320 87656: contig of 2337 bp in length
 * 87657 87756: gap of unknown length
 * 87757 89703: contig of 1947 bp in length
 * 89704 89803: gap of unknown length
 * 89804 91300: contig of 1497 bp in length
 * 91301 91400: gap of unknown length
 * 91401 92768: contig of 1368 bp in length

* 92769 92868: gap of unknown length
 * 92869 94158: contig of 1290 bp in length
 * 94159 94258: gap of unknown length
 * 94259 96077: contig of 1819 bp in length
 * 96078 96177: gap of unknown length
 * 96178 97861: contig of 1684 bp in length
 * 97862 97961: gap of unknown length
 * 97962 99629: contig of 1668 bp in length
 * 99630 99729: gap of unknown length
 * 99730 101901: contig of 2172 bp in length
 * 101902 102001: gap of unknown length
 * 102002 103791: contig of 1790 bp in length
 * 103792 103891: gap of unknown length
 * 103892 105124: contig of 1233 bp in length
 * 105125 105224: gap of unknown length
 * 105225 106434: contig of 1210 bp in length
 * 106435 106534: gap of unknown length
 * 106535 108041: contig of 1507 bp in length
 * 108042 108141: gap of unknown length
 * 108142 109842: contig of 1701 bp in length
 * 109843 109942: gap of unknown length
 * 109943 112159: contig of 2217 bp in length
 * 112160 112259: gap of unknown length
 * 112260 113621: contig of 1362 bp in length
 * 113622 113721: gap of unknown length
 * 113722 115223: contig of 1502 bp in length
 * 115224 115323: gap of unknown length
 * 115324 116353: contig of 1030 bp in length
 * 116354 116453: gap of unknown length
 * 116454 118323: contig of 1870 bp in length
 * 118324 118423: gap of unknown length
 * 118424 119490: contig of 1067 bp in length
 * 119491 119590: gap of unknown length
 * 119591 120620: contig of 1030 bp in length
 * 120621 120720: gap of unknown length
 * 120721 122038: contig of 1318 bp in length
 * 122039 122138: gap of unknown length
 * 122139 123339: contig of 1101 bp in length
 * 123340 123339: gap of unknown length
 * 123340 124595: contig of 1256 bp in length
 * 124596 124695: gap of unknown length
 * 124696 125975: contig of 1280 bp in length
 * 125976 126075: gap of unknown length
 * 126076 127220: contig of 1145 bp in length
 * 127221 127320: gap of unknown length
 * 127321 128724: contig of 1404 bp in length
 * 128725 128824: gap of unknown length
 * 128825 130661: contig of 1837 bp in length
 * 130662 130761: gap of unknown length
 * 130762 132154: contig of 1393 bp in length
 * 132155 132254: gap of unknown length
 * 132255 133282: contig of 1028 bp in length
 * 133283 133382: gap of unknown length
 * 133383 134875: contig of 1493 bp in length
 * 134876 134975: gap of unknown length
 * 134976 136074: contig of 1099 bp in length
 * 136075 136174: gap of unknown length

Alignment Scores:

Pred. No.: 5.67e-33 Length: 143044
 Score: 42.00 Matches: 42
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 21.21% Indels: 0
 DB: 2 Gaps: 0

US-09-966-880A-8 (1-198) x AC094826 (1-143044)

Qy 54 GlyCysHisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGly 73
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 10093 GGCTGCCACGTGGAAATGTTGTTCACGCTACATCTCGGACCTGGACCCGGC 10152

Qy 74 ArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHis 93

Db 10153 CGGTGTACCGGTACCTGCTACCTGCGCCCTGCTACCGACTGTGCGGGCAC 10212
 QY 94 ValAla 95
 Db 10213 GTGGCT 10218
 RESULT 49
 AC120617
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-3215, *** SEQUENCING IN PROGRESS ***,
 60 unordered pieces.
 AC120617
 AC120617.2 GI:21902766
 VERSION HTG: HTGS-PHASE1.
 KEYWORDS Rattus norvegicus.
 SOURCE Rattus norvegicus.
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 177103)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
 Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayale, M., Banks, T.,
 Barbarta, J., Bencon, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buha, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisse, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scher, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Taylor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, P., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.
 Direct Submission
 2 (bases 1 to 177103)
 Worley, K.C.
 Direct Submission
 Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 177103)
 Worley, K.C.
 Direct Submission
 Submitted (23-JUL-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 18, 2002 this sequence version replaced gi:20514478.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GWUN
 Center clone name: CH230-3215
 ----- Summary Statistics
 Sequencing vector: plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 127845 bases at least Q40
 Consensus quality: 134816 bases at least Q30
 Consensus quality: 140379 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 60 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1
 * 1195: contig of 1195 bp in length
 * 1295: gap of unknown length
 * 1296: 2531: contig of 1236 bp in length
 * 2532: 2631: gap of unknown length
 * 2632: 3338: contig of 1307 bp in length
 * 3939: 4038: gap of unknown length
 * 4039: 5140: contig of 1102 bp in length
 * 5141: 5240: gap of unknown length
 * 5241: 6919: contig of 1679 bp in length
 * 6920: 7019: gap of unknown length
 * 7020: 8150: contig of 1131 bp in length
 * 8151: 8250: gap of unknown length
 * 8251: 9652: contig of 1401 bp in length
 * 9652: 11122: contig of 1371 bp in length
 * 9752: 11122: contig of 1371 bp in length
 * 11123: 12225: gap of unknown length
 * 11223: 12525: contig of 1303 bp in length
 * 12526: 12525: gap of unknown length
 * 12626: 14175: contig of 1550 bp in length
 * 14176: 14275: gap of unknown length
 * 14276: 15740: contig of 1465 bp in length
 * 15741: 15840: gap of unknown length
 * 15841: 17373: contig of 1533 bp in length
 * 17374: 17473: gap of unknown length
 * 17474: 20224: contig of 2551 bp in length
 * 20225: 20124: gap of unknown length
 * 20125: 21425: contig of 1301 bp in length
 * 21426: 21525: gap of unknown length
 * 21526: 22828: contig of 1303 bp in length
 * 22829: 23955: gap of unknown length
 * 23956: 24055: contig of 1027 bp in length
 * 24056: 25240: contig of 1185 bp in length
 * 25241: 25340: gap of unknown length
 * 25341: 26933: contig of 1593 bp in length
 * 26934: 27033: gap of unknown length
 * 27034: 28463: contig of 1430 bp in length
 * 28464: 28563: gap of unknown length
 * 28564: 30803: contig of 2240 bp in length
 * 30804: 30903: gap of unknown length
 * 30904: 32330: contig of 1427 bp in length
 * 32331: 32430: gap of unknown length
 * 32431: 34113: contig of 1883 bp in length
 * 34114: 34413: gap of unknown length

* 34414	36200:	contig of 1787	bp in length
* 36201	36300:	gap of unknown	length
* 36301	37867:	contig of 1567	bp in length
* 37868	37967:	gap of unknown	length
* 40374	40374:	contig of 2407	bp in length
* 40375	40474:	gap of unknown	length
* 40475	42090:	contig of 1616	bp in length
* 42091	42190:	gap of unknown	length
* 42191	43912:	contig of 1722	bp in length
* 43913	44012:	gap of unknown	length
* 44013	45650:	contig of 1638	bp in length
* 45651	45750:	gap of unknown	length
* 47521	47922:	contig of 2172	bp in length
* 47923	48022:	gap of unknown	length
* 48023	50032:	contig of 2010	bp in length
* 50033	50132:	gap of unknown	length
* 50133	51867:	contig of 1735	bp in length
* 51868	51967:	gap of unknown	length
* 51968	54813:	contig of 2846	bp in length
* 54814	54913:	gap of unknown	length
* 54914	56337:	contig of 1424	bp in length
* 56338	56437:	gap of unknown	length
* 56438	59133:	contig of 2696	bp in length
* 59134	59233:	gap of unknown	length
* 59234	61478:	contig of 2245	bp in length
* 61479	61578:	gap of unknown	length
* 61579	64321:	contig of 2743	bp in length
* 64322	64421:	gap of unknown	length
* 64422	67033:	contig of 2612	bp in length
* 67034	67133:	gap of unknown	length
* 67134	69227:	contig of 2094	bp in length
* 69228	69327:	gap of unknown	length
* 69328	71941:	contig of 2614	bp in length
* 71942	72041:	gap of unknown	length
* 72042	76107:	contig of 4066	bp in length
* 76108	76207:	gap of unknown	length
* 76208	78643:	contig of 2436	bp in length
* 78644	78743:	gap of unknown	length
* 78744	82956:	contig of 4213	bp in length
* 82957	83056:	gap of unknown	length
* 83057	86461:	contig of 3405	bp in length
* 86462	86561:	gap of unknown	length
* 86562	89125:	contig of 2564	bp in length
* 89126	89225:	gap of unknown	length
* 89226	91743:	contig of 2518	bp in length
* 91744	91843:	gap of unknown	length
* 91844	97444:	contig of 5601	bp in length
* 97445	97544:	gap of unknown	length
* 97445	101058:	contig of 3514	bp in length
* 101059	101158:	gap of unknown	length
* 101159	104772:	contig of 3614	bp in length
* 104773	104872:	gap of unknown	length
* 104873	109395:	contig of 4523	bp in length
* 109396	109495:	gap of unknown	length
* 109496	111919:	contig of 2424	bp in length
* 111920	112019:	gap of unknown	length
* 112020	119114:	contig of 7095	bp in length
* 119115	119214:	gap of unknown	length
* 119215	122620:	contig of 3406	bp in length
* 122621	122720:	gap of unknown	length
* 122721	128279:	contig of 5559	bp in length
* 128280	128379:	gap of unknown	length
* 128380	132063:	contig of 3684	bp in length

Alignment Scores:

Pred. No.:	6.82e-33	Length:	177103
Score:	42.00	Matches:	42
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	21.21%	Indels:	0
DB:	2	Gaps:	0

US-09-966-880A-8 (1-198) x AC120617 (1-177103)

QY	54	GlyCysHisValIcLulLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGly	73
Db	142302	GGCTGCCACGTGAATTTGTTCTACCTACATCTCGAGTGGACCTGGACCCGGC	142361
QY	74	ArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHis	93
Db	142362	CGGTGTTACCGTGTACCTGGTTCACCTTCTGAGCCCTGCTACGACTGTGCGCGCAC	142421
QY	94	ValAla	95
Db	142422	GTGGCT	142427
RESULT	50		
AC119975			
LOCUS		50269 bp	DNA linear HTG 02-MAY-2002
DEFINITION		Mus musculus clone RP24-483K3, LOW-PASS SEQUENCE SAMPLING.	
ACCESSION		AC119975	
VERSION		AC119975.1	GI:20389635
KEYWORDS		HTG; HTGS_PHASE0.	
SOURCE		Mus musculus.	
ORGANISM		Mus musculus	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE		1 (bases 1 to 50269)	
AUTHORS		Birren,B., Linton,L., Nusbaum,C. and Lander,E.	
TITLE		Mus musculus, clone RP24-483K3	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 50269)	
AUTHORS		Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,	
		Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,	
		Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,	
		Chazaro,B., Choepel,X., Colangelo,M., Collins,S., Collymore,A.,	
		Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,	
		Faroo,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,	
		Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,	
		Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,	
		Kamat,A., Karatas,A., Kells,C., Lacombe,K., Lamazares,R.,	
		Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,	
		McLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,	
		McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,	
		Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,	
		Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,	
		Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,	
		Ramond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,	
		Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,	
		Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,	
		Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,	
		Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,	
		Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,	
		Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.	
		Direct Submission	
TITLE		Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome	
JOURNAL		Research, 320 Charles Street, Cambridge, MA 02141, USA	
COMMENT		All repeats were identified using RepeatMasker:	
		Smit, A.F.A. & Green, P. (1996-1997)	
		http://ftp.genome.washington.edu/RM/RepeatMasker.html	
		----- Genome Center	
		Center: Whitehead Institute/ MIT Center for Genome Research	
		Center code: WIBR	
		Web site: http://www-seq.wi.mit.edu	
		Contact: sequence_submissions@genome.wi.mit.edu	
		----- Project Information	
		Center project name: L25744	
		Center clone name: 483_K_3	

- * NOTE: This record contains 65 individual
- * sequencing reads that have not been assembled into
- * contigs. Runs of N are used to separate the reads
- * and the order in which they appear is completely
- * arbitrary. Low-pass sequence sampling is useful for
- * identifying clones that may be gene-rich and allows
- * overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 * 685 784: contig of 684 bp in length
 * 785 1439: contig of 655 bp in length
 * 1440 1539: gap of 100 bp
 * 1540 2214: contig of 675 bp in length
 * 2215 2314: gap of 100 bp
 * 2315 3000: contig of 686 bp in length
 * 3001 3100: gap of 100 bp
 * 3101 3794: contig of 694 bp in length
 * 3795 3894: gap of 100 bp
 * 3895 4576: contig of 682 bp in length
 * 4577 4676: gap of 100 bp
 * 4677 5361: contig of 685 bp in length
 * 5362 5461: gap of 100 bp
 * 5462 6138: contig of 677 bp in length
 * 6139 6238: gap of 100 bp
 * 6239 6900: contig of 662 bp in length
 * 6901 7000: gap of 100 bp
 * 7001 7674: contig of 674 bp in length
 * 7675 7774: gap of 100 bp
 * 7775 8446: contig of 672 bp in length
 * 8447 8546: gap of 100 bp
 * 8547 9219: contig of 673 bp in length
 * 9220 9319: gap of 100 bp
 * 9320 9962: contig of 643 bp in length
 * 9963 10062: gap of 100 bp
 * 10063 10738: contig of 676 bp in length
 * 10739 10838: gap of 100 bp
 * 10839 11520: contig of 682 bp in length
 * 11521 11620: gap of 100 bp
 * 11621 12311: contig of 691 bp in length
 * 12312 12411: gap of 100 bp
 * 12412 13105: contig of 694 bp in length
 * 13106 13205: gap of 100 bp
 * 13206 13865: contig of 660 bp in length
 * 13866 13965: gap of 100 bp
 * 13966 14653: contig of 688 bp in length
 * 14654 14753: gap of 100 bp
 * 14754 15411: contig of 658 bp in length
 * 15412 15511: gap of 100 bp
 * 15512 16182: contig of 671 bp in length
 * 16183 16282: gap of 100 bp
 * 16283 16971: contig of 689 bp in length
 * 16972 17071: gap of 100 bp
 * 17072 17754: contig of 683 bp in length
 * 17755 17854: gap of 100 bp
 * 17855 18535: contig of 681 bp in length
 * 18536 18635: gap of 100 bp
 * 18636 19317: contig of 682 bp in length
 * 19318 19417: gap of 100 bp
 * 19418 20096: contig of 679 bp in length
 * 20097 20196: gap of 100 bp
 * 20197 20879: contig of 683 bp in length
 * 20880 20979: gap of 100 bp
 * 20980 21667: contig of 688 bp in length
 * 21668 21767: gap of 100 bp
 * 21768 22437: contig of 670 bp in length
 * 22438 22537: gap of 100 bp
 * 22538 23197: contig of 660 bp in length
 * 23198 23297: gap of 100 bp
 * 23298 23974: contig of 677 bp in length
 * 23975 24074: gap of 100 bp
 * 24075 24745: contig of 671 bp in length
 * 24746 24845: gap of 100 bp
 * 24846 25522: contig of 677 bp in length
 * 25523 25622: gap of 100 bp
 * 25623 26294: contig of 672 bp in length
 * 26295 26394: gap of 100 bp
 * 26395 27061: contig of 667 bp in length

* 27062 27161: gap of 100 bp
 * 27162 27827: contig of 666 bp in length
 * 27828 27927: gap of 100 bp
 * 27928 28611: contig of 684 bp in length
 * 28612 28711: gap of 100 bp
 * 28712 29395: contig of 684 bp in length
 * 29396 29495: gap of 100 bp
 * 29496 30166: contig of 671 bp in length
 * 30167 30266: gap of 100 bp
 * 30267 30946: contig of 680 bp in length
 * 30947 31046: gap of 100 bp
 * 31047 31700: contig of 654 bp in length
 * 31701 31800: gap of 100 bp
 * 31801 32455: contig of 655 bp in length
 * 32456 32555: gap of 100 bp
 * 32556 33224: contig of 669 bp in length
 * 33225 33324: gap of 100 bp
 * 33325 34012: contig of 688 bp in length
 * 34013 34112: gap of 100 bp
 * 34113 34756: contig of 644 bp in length
 * 34757 34856: gap of 100 bp
 * 34857 35514: contig of 658 bp in length
 * 35515 35614: gap of 100 bp
 * 35615 36297: contig of 683 bp in length
 * 36298 36397: gap of 100 bp
 * 36398 37066: contig of 669 bp in length
 * 37067 37166: gap of 100 bp
 * 37167 37848: contig of 682 bp in length
 * 37849 37948: gap of 100 bp
 * 37949 38606: contig of 658 bp in length
 * 38607 38706: gap of 100 bp
 * 38707 39385: contig of 679 bp in length
 * 39386 39485: gap of 100 bp
 * 39486 40178: contig of 693 bp in length
 * 40179 40278: gap of 100 bp
 * 40279 40935: contig of 657 bp in length
 * 40936 41035: gap of 100 bp
 * 41036 41718: contig of 683 bp in length
 * 41719 41818: gap of 100 bp
 * 41819 42492: contig of 674 bp in length
 * 42493 42592: gap of 100 bp
 * 42593 43275: contig of 683 bp in length
 * 43276 43375: gap of 100 bp
 * 43376 44038: contig of 663 bp in length
 * 44039 44138: gap of 100 bp
 * 44139 44821: contig of 683 bp in length
 * 44822 44921: gap of 100 bp
 * 44922 45602: contig of 681 bp in length
 * 45603 45702: gap of 100 bp
 * 45703 46382: contig of 680 bp in length
 * 46383 46482: gap of 100 bp
 * 46483 47153: contig of 671 bp in length
 * 47154 47253: gap of 100 bp
 * 47254 47933: contig of 680 bp in length
 * 47934 48033: gap of 100 bp
 * 48034 48705: contig of 672 bp in length
 * 48706 48805: gap of 100 bp
 * 48806 49481: contig of 676 bp in length
 * 49482 49581: gap of 100 bp
 * 49582 50269: contig of 688 bp in length.

FEATURES
 Source

1. .50269
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP24-483K3"

BASE COUNT 12034 a 9403 c 9549 g 12707 t 6576 others
 ORIGIN

Alignment Scores:

Pred. No.: 4.09e-29 Length: 50269
 Score: 38.00 Matches: 38
 Percent Similarity: 100.00% Conservative: 0

26130 26229: gap of unknown length
 26230 27842: contig of 1613 bp in length
 27843 27942: gap of unknown length
 27943 29436: contig of 1494 bp in length
 29437 29536: gap of unknown length
 29537 31686: contig of 2150 bp in length
 31687 31786: gap of unknown length
 31787 33456: contig of 1670 bp in length
 33457 33556: gap of unknown length
 33557 34636: contig of 1080 bp in length
 34637 34736: gap of unknown length
 34737 37301: contig of 2565 bp in length
 37302 37401: gap of unknown length
 37402 38699: contig of 1298 bp in length
 38700 38799: gap of unknown length
 38800 40740: contig of 1941 bp in length
 40741 40840: gap of unknown length
 40841 43152: contig of 2312 bp in length
 43153 43252: gap of unknown length
 43253 45781: contig of 2529 bp in length
 45782 45881: gap of unknown length
 45882 48271: contig of 2390 bp in length
 48272 48371: gap of unknown length
 48372 49668: contig of 1597 bp in length
 49669 50068: gap of unknown length
 50069 52282: contig of 2214 bp in length
 52283 52382: gap of unknown length
 52383 54977: contig of 2595 bp in length
 54978 55077: gap of unknown length
 55078 57554: contig of 2577 bp in length
 57555 57754: gap of unknown length
 57755 59700: contig of 2116 bp in length
 59701 59970: gap of unknown length
 59971 62400: contig of 2430 bp in length
 62401 62500: gap of unknown length
 62501 64476: contig of 1976 bp in length
 64477 64576: gap of unknown length
 64577 67748: contig of 3172 bp in length
 67749 67848: gap of unknown length
 67849 70585: contig of 2737 bp in length
 70586 70685: gap of unknown length
 70686 73046: contig of 2361 bp in length
 73047 73146: gap of unknown length
 73147 76312: contig of 3166 bp in length
 76313 76412: gap of unknown length
 76413 79506: contig of 3094 bp in length
 79507 79606: gap of unknown length
 79607 83285: contig of 3679 bp in length
 83286 83385: gap of unknown length
 83386 88188: contig of 4803 bp in length
 88189 88288: gap of unknown length
 88289 93249: contig of 4961 bp in length
 93250 93349: gap of unknown length
 93350 98161: contig of 4812 bp in length.

FEATURES

source
 i..98161
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-98A4"
 BASE COUNT 25850 a 20396 c 20306 g 25999 t 5610 others
 ORIGIN

Alignment Scores:

Pred. No.: 1.3e-24 Length: 98161
 Score: 34.00 Matches: 34
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.17% Indels: 0
 DB: 2 Gaps: 0

US-09-966-880A-8 (1-198) x AC109119 (1-98161)

QY 54 GlyCysHisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGly 73

Db 4378 GGCTCCACGTGGAATTTGTTCTACGCTACATCTCGGACTGGACCTGACCCCGGC 4319
 QY 74 ArgCysTyrArgValThrTrpPheThrSerTrpSerProCys 87
 Db 4318 CGGTGTTACCGTGTACCTGGTTCACTTCTCTGGAGCCCTGC 4277
 RESULT 52
 AC119975/c 50269 bp DNA linear HTG 02-MAY-2002
 Locus Mus musculus clone RP24-483K3, LOW-PASS SEQUENCE SAMPLING.
 DEFINITION AC119975
 ACCESSION AC119975
 VERSION AC119975.1 GI:20389635
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE Mus musculus.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Birren.B., Linton.L., Nusbaum.C. and Lander.E.
 1 (bases 1 to 50269)
 Mus musculus, clone RP24-483K3
 Unpublished
 2 (bases 1 to 50269)
 Birren.B., Linton.L., Nusbaum.C., Lander.E., Ali.A., Allen.N.,
 Anderson.S., Barna.N., Bastien.V., Bloom.T., Boguslavsky.I.,
 Boukhgalter.B., Brown.A., Camarata.J., Campopiano.A., Chang.J.,
 Chazaro.B., Choepel.Y., Colangelo.M., Collins.S., Collymore.A.,
 Cook.A., Cooke.P., DeAtellano.K., Dewar.K., Diaz.J.S., Dodge.S.,
 Fato.S., Ferreira.P., FitzHugh.W., Gage.D., Galagan.J., Gardyna.S.,
 Ginde.S., Gord.S., Goyette.M., Graham.L., Grand-Pierre.N.,
 Hagos.B., Horton.L., Hulme.W., Iliev.I., Johnson.R., Jones.C.,
 Kamat.A., Karatas.A., Kells.C., LaRoque.K., Lamazars.R.,
 Landers.T., Lehoczyk.J., Levine.R., Lindblad-Toh.K., Liu.G.,
 MacLean.C., MacDonald.P., Major.J., Marquis.N., Matthews.C.,
 McCarthy.M., McEwan.P., McKernan.K., Meldrum.J., Menees.L.,
 Mihova.T., Mlenga.V., Murphy.T., Naylor.J., Nguyen.C., Nicol.R.,
 Norbu.C., Norman.C.H., O'Connor.T., O'Donnell.P., O'Neill.D.,
 Oliver.J., Peterson.K., Phunkhang.P., Pierre.N., Pollara.V.,
 Raymond.C., Retta.R., Rieback.M., Riley.R., Rise.C., Rogov.P.,
 Roman.J., Rosetti.M., Roy.A., Santos.R., Schauer.S., Schupbach.R.,
 Seaman.S., Severy.P., Spencer.B., Stange-Thomann.N., Stojanovic.N.,
 Strauss.N., Subramanian.A., Talamas.J., Tesfaye.S., Theodorou.J.,
 Topham.K., Travers.M., Travis.N., Triggillo.J., Vassiliev.H.,
 Viel.R., Vo.A., Wilson.B., Wu.X., Wyman.D., Ye.W.J., Young.G.,
 Zainoun.J., Zembek.L., Zimmer.A. and Zody.M.
 Direct Submission
 Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L25744
 Center clone name: 483_K_3

 * NOTE: This record contains 65 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 * 1 584: contig of 684 bp in length
 * 685 784: gap of 100 bp

```

* 785 1439: contig of 655 bp in length
* 1440 1539: gap of 100 bp
* 1540 2214: contig of 675 bp in length
* 2215 2314: gap of 100 bp
* 2315 3000: contig of 686 bp in length
* 3001 3100: gap of 100 bp
* 3101 3794: contig of 694 bp in length
* 3795 3894: gap of 100 bp
* 3895 4576: contig of 682 bp in length
* 4577 4676: gap of 100 bp
* 4677 5361: contig of 685 bp in length
* 5362 5461: gap of 100 bp
* 5462 6138: contig of 677 bp in length
* 6139 6238: gap of 100 bp
* 6239 6900: contig of 662 bp in length
* 6901 7000: gap of 100 bp
* 7001 7674: contig of 674 bp in length
* 7675 7774: gap of 100 bp
* 7775 8446: contig of 672 bp in length
* 8447 8546: gap of 100 bp
* 8547 9219: contig of 673 bp in length
* 9220 9319: gap of 100 bp
* 9320 9962: contig of 643 bp in length
* 9963 10062: gap of 100 bp
* 10063 10738: contig of 676 bp in length
* 10739 10838: gap of 100 bp
* 10839 11520: contig of 682 bp in length
* 11521 11620: gap of 100 bp
* 11621 12311: contig of 691 bp in length
* 12312 12411: gap of 100 bp
* 12412 13105: contig of 694 bp in length
* 13106 13205: gap of 100 bp
* 13206 13865: contig of 660 bp in length
* 13866 13965: gap of 100 bp
* 13966 14653: contig of 688 bp in length
* 14654 14753: gap of 100 bp
* 14754 15411: contig of 658 bp in length
* 15412 15511: gap of 100 bp
* 15512 16182: contig of 671 bp in length
* 16183 16282: gap of 100 bp
* 16283 16971: contig of 689 bp in length
* 16972 17071: gap of 100 bp
* 17072 17754: contig of 683 bp in length
* 17755 17854: gap of 100 bp
* 17855 18535: contig of 681 bp in length
* 18536 18635: gap of 100 bp
* 18636 19317: contig of 682 bp in length
* 19318 19417: gap of 100 bp
* 19418 20096: contig of 679 bp in length
* 20097 20196: gap of 100 bp
* 20197 20879: contig of 683 bp in length
* 20880 20979: gap of 100 bp
* 20980 21667: contig of 688 bp in length
* 21668 21767: gap of 100 bp
* 21768 22437: contig of 670 bp in length
* 22438 22537: gap of 100 bp
* 22538 23197: contig of 660 bp in length
* 23198 23297: gap of 100 bp
* 23298 23974: contig of 677 bp in length
* 23975 24074: gap of 100 bp
* 24075 24745: contig of 671 bp in length
* 24746 24845: gap of 100 bp
* 24846 25522: contig of 677 bp in length
* 25523 25622: gap of 100 bp
* 25623 26294: contig of 672 bp in length
* 26295 26394: gap of 100 bp
* 26395 27061: contig of 667 bp in length
* 27062 27161: gap of 100 bp
* 27162 27827: contig of 666 bp in length
* 27828 27927: gap of 100 bp
* 27928 28611: contig of 684 bp in length
* 28612 28711: gap of 100 bp
* 28712 29395: contig of 684 bp in length

```

```

* 29396 29495: gap of 100 bp
* 29496 30166: contig of 671 bp in length
* 30167 30266: gap of 100 bp
* 30267 30946: contig of 680 bp in length
* 30947 31046: gap of 100 bp
* 31047 31700: contig of 654 bp in length
* 31701 31800: gap of 100 bp
* 31801 32455: contig of 655 bp in length
* 32456 32555: gap of 100 bp
* 32556 33224: contig of 669 bp in length
* 33225 33324: gap of 100 bp
* 33325 34012: contig of 688 bp in length
* 34013 34112: gap of 100 bp
* 34113 34756: contig of 644 bp in length
* 34757 34856: gap of 100 bp
* 34857 35514: contig of 658 bp in length
* 35515 36297: contig of 683 bp in length
* 36298 36397: gap of 100 bp
* 36398 37066: contig of 669 bp in length
* 37067 37166: gap of 100 bp
* 37167 37848: contig of 682 bp in length
* 37849 37948: gap of 100 bp
* 37949 38606: contig of 658 bp in length
* 38607 38706: gap of 100 bp
* 38707 39385: contig of 679 bp in length
* 39386 39485: gap of 100 bp
* 39486 40178: contig of 693 bp in length
* 40179 40278: gap of 100 bp
* 40279 40935: contig of 657 bp in length
* 40936 41035: gap of 100 bp
* 41036 41718: contig of 683 bp in length
* 41719 41818: gap of 100 bp
* 41819 42492: contig of 674 bp in length
* 42493 42592: gap of 100 bp
* 42593 43275: contig of 683 bp in length
* 43276 43375: gap of 100 bp
* 43376 44038: contig of 663 bp in length
* 44039 44138: gap of 100 bp
* 44139 44821: contig of 683 bp in length
* 44822 44921: gap of 100 bp
* 44922 45602: contig of 681 bp in length
* 45603 45702: gap of 100 bp
* 45703 46382: contig of 680 bp in length
* 46383 46482: gap of 100 bp
* 46483 47153: contig of 671 bp in length
* 47154 47253: gap of 100 bp
* 47254 47933: contig of 680 bp in length
* 47934 48033: gap of 100 bp
* 48034 48705: contig of 672 bp in length
* 48706 48805: gap of 100 bp
* 48806 49481: contig of 676 bp in length
* 49482 49581: gap of 100 bp
* 49582 50269: contig of 688 bp in length.

FEATURES
    Location/Qualifiers
        1..50269
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="RP24-483K3"
            /clone_lib="RPCI-24 Male Mouse BAC"
BASE COUNT 12034 a 9403 c 9549 g 12707 t 6576 others
ORIGIN

Alignment Scores:
Pred. No.: 9.83e-06 Length: 50269
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.08% Indels: 0
Db: 2 Gaps: 0

US-09-966-880A-8 (1-198) x AC119975 (1-50269)

```

```

QY      26 GluThrTyrLeuCystrValVallysArgArgAspSerAlaThrSer 41
      |||||||
Db      35893 GAGACCTACCTCTGCTACGTGTGAAGAGAGAGATAGTCCACCTCC 35846

RESULT 53
AC068309/c 67244 bp DNA linear HTG 16-OCT-2001
LOCUS      Mus musculus chromosome 13 clone RP23-298J20, *** SEQUENCING IN
DEFINITION PROGRESS ***, 22 unordered pieces.
ACCESSION AC068309
VERSION    3 GI:16118088
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     Mus musculus.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 67244)
Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
Gorell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Loulseghe,H., Lozado,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogih,M., Parish,B.,
Perzel,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
Worley,K. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 67244)
Worley,K.C.
Direct Submission
Submitted (02-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 14, 2001 this sequence version replaced gi:11094637.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MADZ
Center clone name: RP23-298J20
----- Summary Statistics
Sequencing vector: M13: L08821
Chemistry: Dye-Primer Bodipy: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 69254 bases at least Q40
Consensus quality: 93836 bases at least Q30
Consensus quality: 103881 bases at least Q20
Estimated insert size: 89529: sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-ef estimation
Quality coverage: 1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4843: contig of 4843 bp in length
* 4844 4943: gap of unknown length
* 4944 9846: contig of 4903 bp in length

```

```

* 9847 9946: gap of unknown length
* 9947 15031: contig of 5085 bp in length
* 15032 15131: gap of unknown length
* 15132 18316: contig of 3185 bp in length
* 18317 18416: gap of unknown length
* 21312 21317: contig of 2896 bp in length
* 21412 21417: gap of unknown length
* 25595 25595: contig of 4183 bp in length
* 25596 25595: gap of unknown length
* 25696 28983: contig of 3288 bp in length
* 28984 31371: contig of 2288 bp in length
* 31372 31471: gap of unknown length
* 31472 34502: contig of 3031 bp in length
* 34503 34603: gap of unknown length
* 34603 37164: contig of 2562 bp in length
* 37165 37264: gap of unknown length
* 37265 40058: contig of 2794 bp in length
* 40059 40158: gap of unknown length
* 40159 42440: contig of 2282 bp in length
* 42441 42540: gap of unknown length
* 42541 44901: contig of 2361 bp in length
* 44902 45002: gap of unknown length
* 45003 47782: contig of 2781 bp in length
* 47783 50069: contig of 2187 bp in length
* 50070 50169: gap of unknown length
* 50170 52502: contig of 2333 bp in length
* 52503 52603: gap of unknown length
* 52603 55087: contig of 2485 bp in length
* 55088 55187: gap of unknown length
* 55188 57301: contig of 2114 bp in length
* 57302 57401: gap of unknown length
* 57402 59994: contig of 2593 bp in length
* 59995 60095: gap of unknown length
* 60095 62360: contig of 2266 bp in length
* 62361 62460: gap of unknown length
* 62461 65125: contig of 2665 bp in length
* 65126 65225: gap of unknown length
* 65226 67244: contig of 2019 bp in length.

FEATURES             Location/Qualifiers
     1..67244
         /organism="Mus musculus"
         /db_xref="taxon:10090"
         /chromosome="13"
         /clone="RP23-298J20"

BASE COUNT  17937 a 13995 c 14816 g 18342 t 2154 others
ORIGIN

Alignment Scores:
Pred. No.:      30.1      Length:      67244
Score:          10.00     Matches:      10
Percent Similarity: 100.00%  Conservat:    0
Best Local Similarity: 100.00%  Mismatches:   0
Query Match:      5.05%     Indels:      0
DB:              2         Gaps:      0

US-09-966-880A-8 (1-198) x AC068309 (1-67244)

QY      169 SerValArgLeuSerArgGlnLeuArgArg 178
      |||||||
Db      12769 TCTGTGAGGCTGAGCAGCAGCTTCGTAGA 12740

RESULT 54
DMSPALTR
LOCUS      D.melanogaster spalt-related gene.
DEFINITION Y07653
ACCESSION  Y07653
VERSION     Y07653.1 GI:1684680
KEYWORDS    spalt gene; zinc finger.
SOURCE      Drosophila melanogaster.
ORGANISM    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

```



```

GQSKPDDPYDEASSKRKARCDVCMALDHDFGNEFSLAGHDSRGSYAFRTAMDHPAR
VAKLAIDGVFLLEALEHCNARFATQWWHFEFFAQHDKPEQAI SADPLRWYGGTEDLM
GKNYAVFLAQVQHSITQGMGLDYRAGVRIDHLDHEDRNSGKLOCPLRLVLSLRD
DLEDLYNDVLAIWQPAELPVSGIGDSGHMAEAPLEVARELKSFFLA"
complement(1637. .2269)
/gene="AGR_pat_729"
CDS
/gene="AGR_pat_729"
complement(1637. .2269)
/gene="AGR_pat_729"
/translation="putative tetra family transcriptional
regulator"
/codon_start=1
/transl_table=11
/product="AGR_pat_729p"
/protein_id="AAK90871.1"
/db_xref="GI:15162312"
/translation="MTCDSHNSWSALHVPKERTQPLSEAPLADARNREKLIE
VASAFESQGAASLEADTARAGVGIGTLYRHFTREHLVESYRRELELATAARDL
IAEPEAPAEIEMNRYFNMAKRGMSKLKLFSTNSITFAEGSKLHTFAEGLLD
NAVADAVKGIDIEADVLNALFGVSIPEGPEWRDRAHVRVLVMDGLRR"
2444. .3433
/gene="AGR_pat_731"
CDS
2444. .3433
/gene="AGR_pat_731"
/translation="HYPOTHETICAL 35.0 KD PROTEIN IN INTF-EAEH
INTERGENIC REGION"
/codon_start=1
/transl_table=11
/product="AGR_pat_731p"
/protein_id="AAK90872.1"
/db_xref="GI:15162313"
/translation="MTCDSADIRHNGAPRWLPNADIPVAVPQKAFSTDTACDIIS
FAIANIAGLVGCTVIEIRGSSRPLGAHMAVQDRCYGVSGGCTEAAVAEAL
QAKSGHDFVLLGEGSPFDVLPKGGGIVATHLLKDVDPCLCSVNLVDRGSATL
QCYGLELSGSCPGATGWDGDFLQSYRKPSPVIEGSRLEVETAVNAKAGYEV
IFEDHTAGDSFDSLIDATVAILLHDLLEMPILGALVATPFYIGALGSSRTHKE
RYERURQIGHSEEAIRAKPIGLFGRADAQSLAFSVVADIAASRSCVMHS"
3561. .4094
/gene="AGR_pat_732"
CDS
3561. .4094
/gene="AGR_pat_732"
/translation="Oxidoreductase, iron-sulfur subunit - Deinococcus
radiodurans (strain R1)"
/codon_start=1
/transl_table=11
/product="AGR_pat_732p"
/protein_id="AAK90873.1"
/db_xref="GI:15162314"
/translation="MOTINGTAVEADINTSLDLMRNVLGTGTGKGCDCQACGA
CTVLVDGERINSLALAVQYGRSITTVIEGLAANGNLHPLOKAFIEHDGQCYCTPG
QLCSAIGMAGEIERNIPSVVTADLSNSITVDDTEIRERMSGNLCRCGAYNGIVEATS
DTLVQQTAVPSEERARA"
4091. .5080
/gene="AGR_pat_735"
CDS
4091. .5080
/gene="AGR_pat_735"
/translation="Oxidoreductase - Deinococcus radiodurans (strain
R1)"
/codon_start=1
/transl_table=11
/product="AGR_pat_735p"
/protein_id="AAK90874.1"
/db_xref="GI:15162315"
/translation="MNIIFSYPQSDTSAIALKQAGTAKYLAGGTNLVDLMRETVER
PETLVDTHTLSSITERTDGLSVAGVKNALAAHDAVRAOPFLLSRAILAGASAOL
RNATVGNILQTRCRVFDYNAARCNKRONPSGCDALGEGNRYHAILGASECVATH
LSPDCVALVDASVHLQGPAGTRSLPLRDLHRLPGHTPEIETELDODLLITAVEIPS
LDFKGRQYRVRDASTAFVLSAALVDVDDAGNVCRLALGGVVAHPWRALKAE
SLRKQQAESAESFAAEALADAILQDNFAKTELAKRTIVAVLEELKGESA"
5077. .7380
/gene="AGR_pat_739"
CDS
5077. .7380
/gene="AGR_pat_739"
/translation="probable oxidoreductase (EC 1.-.-.-) DRA0237

```

```

(similarity) - Deinococcus radiodurans (strain R1)"
/codon_start=1
/transl_table=11
/product="AGR_pat_739p"
/protein_id="AAK90875.1"
/db_xref="GI:15162316"
/translation="MSTDOSOTLEAGKLGWQPAVTSDDLRLKIHGALGQGLSRLEGPM
KVQCTRFBAEFPYDNIYSYALAFSTVARGVRIDLVNSAAEAAGVILVYTHKNAPRM
KAPLSMSSPTAGASNLVPMQSDIEHNGQVIALVIAETOEADHAASLVTKAYELL
PAVTSFDDAKKSPHOLDNLJGQPPFIDIGDAETALANAEXKVDVLYVTRPRHNAIEL
HAATVVMKDDLRVDSQLDLTLTQGLADIFGLDVSXVHVTSYVGGGFGGKCFMDH
QILACAAARLAGRPVRLMISREGVFRIGGRTVTEQORVALCAKADGTLDALIHTGTAA
MILHNSCEQOFTFPAHLYAAKTFPHIGQDVADMMLANTFMRAFGESVGTFALECALD
ELAEKGLDPIELRRRIEPEKDPDTTGKPFSSRYLLEAYEKGAEOFSGHKSQTPROKR
EGEWLIGMCATATYPIHRFPFGAVRIKLTVDGRVTVSTAVHDMGMDGTATAMQHILAA
RLGLPLDHVTFEYGDLSKPRVITAGSTQSTASIGAVIAATEVIEELJLISGNDSP
AGLSLVEARDGGLSHISDNLSFESQSLRRAGREELVCEAFAPAPAEAFSMHS
YGAQFCVHVSAARTGETRVSRFLGSPDAGQILNSKMATSQFKGII MGICGLATETTN
FDERTGRIVNASLADYHVPQMDVPAIEILYTNKPDQPAQPMGANGICEIGITGVGAAY
ANAVYNATGVVRDLPTLTKLGTGLD"
complement(7724. .8416)
/gene="AGR_pat_742"
CDS
complement(7724. .8416)
/gene="AGR_pat_742"
/translation="conserved hypothetical protein - Deinococcus
radiodurans (strain R1)"
/codon_start=1
/transl_table=11
/product="AGR_pat_742p"
/protein_id="AAK90876.1"
/db_xref="GI:15162317"
/translation="MRNMNVGMMKLTETPSODKAEGRNPVTVAIILAAKASR
MGDGAHKLAEDEGKPLVRSFAIALTADAASVIVYVTHRRSEIEAALLGLGVTTIN
NPDYASGASSLIAGFAPACADGVLVWLADMPGVTIGDLDLILISAPRASGRALV
VYRSGRGRGPVILPRLSLGAVMRLEGDVGARQIETISGLRIIVETIGDSALHDVDT
EAVVAAGGIILT"
complement(8501. .9256)
/gene="AGR_pat_743"
CDS
complement(8501. .9256)
/gene="AGR_pat_743"
/translation="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_pat_743p"
/protein_id="AAK90877.1"
/db_xref="GI:15162318"
/translation="MIKTIATVLTATTTWCATSSNAREGIVRVAVYKGPASCEDC
SETMRKSIMALGPEYSDVYVGAEEKIDISDETLSKFDIYVOPGGGODIPAPFSLGEO
RVEAIRFVAGGGFGSCMGAYLASGGLIDDELESEVGRFPPVSSIISTAVST
SWNGEKQTVFQDQPYMKSEPKDIRGFKKIATYQNGDIAAARYHFGNGIVVLGTGHP
GGSWFHEAGIPIDKMPDHRVLSVMDEFREPR"
complement(9366. .10823)
/gene="AGR_pat_745"
CDS
complement(9366. .10823)
/gene="AGR_pat_745"
/translation="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_pat_745p"
/protein_id="AAK90878.1"
/db_xref="GI:15162319"
/translation="MPSRKLKLSLPLFAALASPLVQOTVAATLIVSISMGQAVAG
SPVRMNPALLGDMRRSLAEAFSEAKGDSKRRSVAVLAAITAOVQSILETAI
AANDASIRATKALSHAVGNLOTRYSLSAKCADVTRALORFNAAEYSSRFLAK
HKASAAATASDVQALRRKRVADLSYRMOOLEDDVADNAALRGEVIRMRREIATOSTIP
ANAYDQRLSFTLTVLSGSPAAALSITTRITYPSYVADPEVRSIRIDTWDGYDGY
YQGVADSYSEPVISSPVSTETRTYQNTYETIYSVTNETTNIYQTMPEEDLSDV
AVVPVQGVNFVEDISANKSGSEVDFESNTYETIYQTMPEEDLSDV
PDVDDQDEPPPAIPSPLEHNSNDNAGTFDDTAGSVSPDHDDKIPWAPEDRISDIHST
PDTLPSSENDEPRAGRVQEDKASPEEYKATP"
2307 a 3145 c 3105 g 2388 t

```

BASE COUNT
ORIGIN

Alignment Scores:

Pred. No.: 72.4 Length: 10945
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.55% Indels: 0
 DB: 1 Gaps: 0

US-09-966-880A-8 (1-198) x AE007918 (1-10945)

QY 125 GYLeuAtgArgLeuHisArgAlaGly 133

Db 3680 GGCTTGGCGCGCTGCACCGTGTGTGT 3706

RESULT 56

AE008970

LOCUS

DEFINITION Agrobacterium tumefaciens str. C58 AT plasmid section 45 of 49 of the complete sequence.

ACCESSION

VERSION AE008970

KEYWORDS

SOURCE AE008970.1 GI:17743952

ORGANISM

Agrobacterium tumefaciens str. C58 (U. Washington).
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.

REFERENCE

AUTHORS 1 (bases 1 to 11013)

Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Zhou, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F., Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Krespan, W., Perry, M., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V. and Nester, E.W.
 The genome of the natural genetic engineer Agrobacterium tumefaciens C58
 Science 294 (5550), 2317-2323 (2001)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS 2 (bases 1 to 11013)

Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Zhou, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F., Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Krespan, W., Perry, M., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V. and Nester, E.W.
 Direct Submission
 Submitted (27-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA 98195-7242, USA

TITLE

JOURNAL

FEATURES

SOURCE

1..11013
 /organism="Agrobacterium tumefaciens str. C58 (U. Washington)"
 /strain="C58"
 /db_xref="taxon:180835"
 /plasmid="At"
 538..1398
 /gene="Atu5495"
 /genes="Atu5495"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=1
 /product="hydrolase"

gene

CDS

/protein_id="AAL46181.1"
 /db_xref="GI:17743953"
 /translation="MFENFRLEMKLPETKLRVRHGGSPPLLLLGHHPRTHTTWYKV
 APIIAEHFTVICPDLRGFGQSKDDPDYDYBASSKRKARDCVALMDHFGFNEFSLAG
 HDRGSYTAFTAMDHPARVAKAILLDGVPILEALEHCNARFATQWHHFFAQHDKPE
 QAISADPLRWYGGTDLGMKNYADFRLAVQDHSITQGMGLGDIYRAGVIRLHLDHREDR
 NSGRKLCPLRLVLSLRDLDLEDLYNDVLAIQWAEPLVPSGLGIDSGHHMAEAPLEY
 ARELKSFFLA"
 complement(1636..2220)
 /gene="Atu5496"
 complement(1636..2220)
 /gene="Atu5496"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="transcriptional regulator, Tetr family"
 /protein_id="AAL46182.1"
 /db_xref="GI:17743954"
 /translation="MPKERTQPPLEAPLRADARRNRREKLIEVASLAFSEQGAASLE
 DIARRAGVGIGTLVRFPTREHLVESVYRRELELLATAARDLIAEPEAEATEENNR
 FVYNAAKRGWANSIKLLFTSNSTLFAESKLLHTAFEGLLDNAVKADAVKGDIEAAD
 VLNALFGIYSIPEGEWRDRAHRIVRLVMDGLRR"
 2443..3432
 /gene="Atu5497"
 2443..3432
 /gene="Atu5497"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAL46183.1"
 /db_xref="GI:17743955"
 /translation="MTQSDADIIRHNGGAPRLPRNADIPYAVPQKAFSTDTACDILS
 FAANKAGLQVGLCTLVEIRGSSRPLGAHMAVAQDGRYCGYSGGCTEAAVAEAL
 QAMKSGHDFRVLLEGSPFFDIVPCGGITVAIHLKDDVPLCVSNMLVDGRRSATL
 CYSGLLESLSGSPGARTGWDGIDFLTYSRPRPVIFSGRLEVTAVNVAKGAGYEY
 IFFDTAGDSFSSLLIDADTAVALLHDLLEMPILGAALVATPFYIGALGSSRTHK
 RVERLROIGHSEETARIKAPIGLFGARDQAQSLAFSVVADIAASRQSCVMHS"
 3560..4093
 /gene="Atu5498"
 3560..4093
 /gene="Atu5498"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="oxidoreductase with iron-sulfur subunit"
 /protein_id="AAL46184.1"
 /db_xref="GI:17743956"
 /translation="MQFTINGPAVEVEADIFTLSLLDMRNYLGFCTKKGCDOGACGA
 CTVLVDGRINSLALAVOYQGRSITTVYEGLAANGNLHPLQKAFIEHGGFCQGYCTPG
 QLCSAIGHAGEIERNIPSVVTADLSSNITVDDTETIRMSGNLRCGAYNGIVEAIS
 DTLVQQTAVPSEERARA"
 4090..5079
 /gene="Atu5499"
 4090..5079
 /gene="Atu5499"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="oxidoreductase"
 /protein_id="AAL46185.1"
 /db_xref="GI:17743957"
 /translation="MNIFSYHQPSDTASAIKQAGPTAKYLAGTNLVDLMRETVER
 PETLVDVTHLSIIERRDTGSLLVAGVKNTALAAHDAVRAQVPLLSRLAGASAOI
 RNVAIVGGNIQRTCRVYFDNAARKRQPSGCDALGFNRVHAIIILCASCSVATH
 PSDMCVALVALDASVHLOGPACTRSLPLRDLHRLPCHTPEIETELDTDLITAVEIPS
 LPFAGRSYRKVRDRASAFALYSVAALDVEDDAGNVKDVRLALGVAHKPWRALKAE
 SRLKQQAESFRAAAEAEALADALGLQDNAPFKIELAKRTIVAVLEELKGESA"
 5076..7379

gene

```
CDS
/feature="Atu5500"
5076..7379
/feature="Atu5500"
/feature="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="dehydrogenase"
/protein_id="AAL46186.1"
/db_xref="GI:17743958"
/translation="MSTDSQTLPEAGKLGKRMQPAVTSDDLRLKHGALQGQISRIEGPM
KYQGRTRFAEPYDNIYSVLAALFSTVARGVRIDLVNLSAAEAPGVILVMTHTKNAPRM
KAPLSMSPPTAAGASNLPMQSDIHWNGOPIALVLAETQDAADAAASLVATKYELL
PAVTSFDDAKSPROLDNLGPPFIDIDGAETALANAQVKVLDVLRPRNNHNAIEL
HAATVYWDDELVRHDAQQLDLTTGQADLIFGLDYSKVHVTSVPYVGGSGKCFWDH
QLIACAARGLRGPVRLMSREGVFRIGGRVTQQRVALGAKAGDTLDLHITGTAA
MTLHNSCEPQFTFPAHLYIAKTFHIGQVADMDMLANTFMRAPESVGTFALECALD
ELAERLGDPIELRPEKDPFTGKPFSSRYLIEAYEKAEQCGWKRSTQPKQR
EGEWLIGMCATATYHRRPGGAVRIKLTVDGRVTSTAVHDMGMGTATAGMOHLAA
RLGLPLDHTFVFGSKLPRGVIAGSGTQTSIGGAVTAATEVLVEELIKLGNDSPL
AGLSLLEVEARDGGLSHLSDNLSFESYQSLRRAGREELVCEAPAEAPAEAFSMHS
YGAQCEVHVSARTGTVRSRFLGSDAGQILNSKMATSFQKGLIMIGLALTEETN
FDQGRIVNASLADYHVPQMDVPAIEILYTNKPDPOAPMGARGIGEIGITGVGAAY
ANAVYNATGVRVRLDPLTKLTGGLD"
complement(7723..8385)
/feature="Atu5501"
complement(7723..8385)
/feature="Atu5501"
/feature="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL46187.1"
/db_xref="GI:17743959"
/translation="MKITRPSQDKAEGNRRPVTVAAILAAGKASRMGDDGAHKL
AFEDKPLVRSAAELADRAASVIVTGHRRSEIAALLGIVGVTNNPDVYASGAS
SLIAGFSAPACDAGDGLVLMADMEFVTIGDLDTLISAFRASGRAIVRAVSRGRGN
PVILPSRLNGAVMRLEGDVGARQIETSLGRIIDVEIGDSAHLDVDTPEAVVAAAGIL
RT"
complement(8500..9255)
/feature="Atu5502"
complement(8500..9255)
/feature="Atu5502"
/feature="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL46188.1"
/db_xref="GI:17743960"
/translation="MIKTIATLTVIATTWCATSSNAREGIVRVAVYRGPAQCEDC
SEFMKRSIMALGPEYSDVYVGAEEKIDIDSETLSEKFDIYVQPGGQDIPAAESLGEQ
RVETLRVFEAGGFFGCMGAYLASGSGIGLIDDELESEVGRGPPVSSIIISTAVST
SWNGEKQTVFFODGPKYMEPKDKIRGFKKIATYONGDIAAARYHFGNGIVLVTGPHEA
GGSWFHEAGIPIDKMPDHRVLSVNDSEFRSEPR"
complement(9365..10822)
/feature="Atu5503"
complement(9365..10822)
/feature="Atu5503"
/feature="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL46189.1"
/db_xref="GI:17743961"
/translation="WPRSKLKLVLTPFAALAASPLVQQTVAATVATVSLSMQAVAG
SPVRNAPALLGDMRSLAEIAFSESEAKGSKRSRGAVALAALTEATQAVOSLETAI
AAHDAASRISRAFKALSHAVGNLQTRYSLSSAKCDVATRALQRFNAWAAYSSREVLAK
```

Alignment Scores:
Pred. No.: 72.8 Length: 11013

```
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.55% Indels: 0
DB: 1 Gaps: 0

US-09-966-880A-8 (1-198) x AE008970 (1-11013)

QY 125 GlyLeuAAGAgLeuHisArgAlaGly 133
Db 3679 GGCATTGCGCGCGCTGCACCGTCTCGT 3705

RESULT 57
AC100984
LOCUS AC100984 Mus musculus clone RP23-76J21, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC100984
ACCESSION AC100984
VERSION AC100984.1 GI:17059758
HTG: HTGS_PHASE0.
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 59915)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-76J21
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 59915)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhallal,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyne,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Jones,B., Heatford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazar,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nuyeh,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,N., Schupbach,K.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trifilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L14683
Center clone name: 76_J_21
-----
* NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful to:
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
```

* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 695: contig of 695 bp in length
* 696 795: gap of 100 bp
* 796 1494: contig of 699 bp in length
* 1495 1594: gap of 100 bp
* 1595 2301: contig of 707 bp in length
* 2302 2401: gap of 100 bp
* 2402 3107: contig of 706 bp in length
* 3108 3207: gap of 100 bp
* 3208 3905: contig of 698 bp in length
* 3906 4005: gap of 100 bp
* 4006 4712: contig of 707 bp in length
* 4713 4812: gap of 100 bp
* 4813 5519: contig of 707 bp in length
* 5520 5619: gap of 100 bp
* 5620 6336: contig of 717 bp in length
* 6337 6436: gap of 100 bp
* 6437 7155: contig of 719 bp in length
* 7156 7255: gap of 100 bp
* 7256 7980: contig of 725 bp in length
* 7981 8080: gap of 100 bp
* 8081 8798: contig of 718 bp in length
* 8799 8998: gap of 100 bp
* 8999 9616: contig of 718 bp in length
* 9617 9716: gap of 100 bp
* 9717 10436: contig of 720 bp in length
* 10437 10536: gap of 100 bp
* 10537 11266: contig of 730 bp in length
* 11267 11366: gap of 100 bp
* 11367 12073: contig of 707 bp in length
* 12074 12173: gap of 100 bp
* 12174 12874: contig of 701 bp in length
* 12875 12974: gap of 100 bp
* 12975 13673: contig of 699 bp in length
* 13674 13773: gap of 100 bp
* 13774 14447: contig of 674 bp in length
* 14448 14547: gap of 100 bp
* 14548 15251: contig of 704 bp in length
* 15252 15351: gap of 100 bp
* 15352 16052: contig of 701 bp in length
* 16053 16152: gap of 100 bp
* 16153 16873: contig of 721 bp in length
* 16874 16973: gap of 100 bp
* 16974 17694: contig of 721 bp in length
* 17695 17794: gap of 100 bp
* 17795 18516: contig of 722 bp in length
* 18517 18616: gap of 100 bp
* 18617 19334: contig of 718 bp in length
* 19335 19434: gap of 100 bp
* 19435 20155: contig of 721 bp in length
* 20156 20255: gap of 100 bp
* 20256 20961: contig of 706 bp in length
* 20962 21061: gap of 100 bp
* 21062 21764: contig of 703 bp in length
* 21765 21864: gap of 100 bp
* 21865 22563: contig of 699 bp in length
* 22564 22663: gap of 100 bp
* 22664 23383: contig of 720 bp in length
* 23384 23483: gap of 100 bp
* 23484 24206: contig of 723 bp in length
* 24207 24306: gap of 100 bp
* 24307 25024: contig of 718 bp in length
* 25025 25124: gap of 100 bp
* 25125 25820: contig of 696 bp in length
* 25821 25920: gap of 100 bp
* 25921 26654: contig of 734 bp in length
* 26655 26754: gap of 100 bp
* 26755 27450: contig of 696 bp in length
* 27451 27550: gap of 100 bp
* 27551 28255: contig of 705 bp in length
* 28256 28355: gap of 100 bp

* 28356 29049: contig of 694 bp in length
* 29050 29149: gap of 100 bp
* 29150 29853: contig of 704 bp in length
* 29854 29953: gap of 100 bp
* 29954 30670: contig of 717 bp in length
* 30671 30770: gap of 100 bp
* 30771 31474: contig of 704 bp in length
* 31475 31574: gap of 100 bp
* 31575 32287: contig of 713 bp in length
* 32288 32387: gap of 100 bp
* 32388 33115: contig of 728 bp in length
* 33116 33215: gap of 100 bp
* 33216 33925: contig of 710 bp in length
* 33926 34025: gap of 100 bp
* 34026 34748: contig of 723 bp in length
* 34749 34848: gap of 100 bp
* 34849 35542: contig of 694 bp in length
* 35543 35642: gap of 100 bp
* 35643 36326: contig of 684 bp in length
* 36327 36426: gap of 100 bp
* 36427 37132: contig of 706 bp in length
* 37133 37232: gap of 100 bp
* 37233 37955: contig of 723 bp in length
* 37956 38055: gap of 100 bp
* 38056 38768: contig of 713 bp in length
* 38769 38868: gap of 100 bp
* 38869 39572: contig of 704 bp in length
* 39573 39672: gap of 100 bp
* 39673 40401: contig of 729 bp in length
* 40402 40501: gap of 100 bp
* 40502 41211: contig of 710 bp in length
* 41212 41311: gap of 100 bp
* 41312 42044: contig of 733 bp in length
* 42045 42144: gap of 100 bp
* 42145 42846: contig of 702 bp in length
* 42847 42946: gap of 100 bp
* 42947 43655: contig of 709 bp in length
* 43656 43755: gap of 100 bp
* 43756 44459: contig of 704 bp in length
* 44460 44559: gap of 100 bp
* 44560 45262: contig of 703 bp in length
* 45263 45362: gap of 100 bp
* 45363 46066: contig of 704 bp in length
* 46067 46166: gap of 100 bp
* 46167 46887: contig of 721 bp in length
* 46888 46987: gap of 100 bp
* 46988 47708: contig of 721 bp in length
* 47709 47808: gap of 100 bp
* 47809 48509: contig of 701 bp in length
* 48510 48609: gap of 100 bp
* 48610 49333: contig of 724 bp in length
* 49334 49433: gap of 100 bp
* 49434 50152: contig of 719 bp in length
* 50153 50252: gap of 100 bp
* 50253 51065: contig of 813 bp in length
* 51066 51165: gap of 100 bp
* 51166 51842: contig of 677 bp in length
* 51843 51942: gap of 100 bp
* 51943 52652: contig of 710 bp in length
* 52653 52752: gap of 100 bp
* 52753 53461: contig of 709 bp in length
* 53462 53561: gap of 100 bp
* 53562 54264: contig of 703 bp in length
* 54265 54364: gap of 100 bp
* 54365 55054: contig of 690 bp in length
* 55055 55154: gap of 100 bp
* 55155 55882: contig of 728 bp in length
* 55883 55982: gap of 100 bp

Alignment Scores:

Pred. No.: 315 Length: 59915
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0

```

Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 4.55%                 Indels: 0
DB: 2                               Gaps: 0

US-09-966-880A-8 (1-198) x AC100984 (1-59915)

Qy 58 GluLeuLeuPheLeuArgTyrlSer 66
   |||||
Db 51017 GAATTGTTATTTTGAGATACATCTCC 51043

RESULT 58
AC014357/c
LOCUS AC014357 103746 bp DNA linear HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
           pieces.
ACCESSION AC014357
VERSION AC014357.1 GI:6436978
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 103746)
Adams,M. and Venter,J.C.
Direct Submission
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10212654 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
     source              1..103746
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"

BASE COUNT  29847 a 21798 c 21974 g 30127 t
ORIGIN
117 GluAspArgLysAlaGluProGluGly 125
|||||
55591 GAGACAGACAGGCGAGCCAGAGCGT 55565

Alignment Scores:
Pred. No.: 506          Length: 103746
Score: 9.00            Matches: 9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 4.55%              Indels: 0
DB: 2                      Gaps: 0

US-09-966-880A-8 (1-198) x AC014357 (1-103746)

Qy 117 GluAspArgLysAlaGluProGluGly 125
   |||||
Db 55591 GAGACAGACAGGCGAGCCAGAGCGT 55565

RESULT 59
AC087457
LOCUS AC087457 141899 bp DNA linear PRI 30-NOV-2001
DEFINITION Homo sapiens chromosome 15, clone RP11-814P5, complete sequence.
ACCESSION AC087457
VERSION AC087457.5 GI:17155075
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 141899)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Unpublished
JOURNAL Homo sapiens chromosome 15, clone RP11-814P5
2 (bases 1 to 141899)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,
Camatara,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Furo,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meidrin,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nanyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Riser,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Submitted (30-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 141899)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camatara,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Furo,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Grand-Pierre,N.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meidrin,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nanyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Riser,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Submitted (30-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 141899)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camatara,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Furo,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Grand-Pierre,N.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meidrin,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nanyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Riser,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Submitted (30-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Nov 30, 2001 this sequence version replaced gi:15528911.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 Only the first 141.9 kilobases of this clone are being submitted.
 The remainder overlaps accession number AC018868.
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L11876
 Center clone name: 814_P_5

FEATURES

source Location/Qualifiers
 1..141899
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15"
 /clone="RP11-814P5"
 /clone_lib="RPC1-11 Human Male BAC"
 1..274
 /note="single clone coverage"
 1..15
 /note"<30 qual single clone coverage"
 30..35
 /note"<30 qual single clone coverage"
 293..298
 /note"<30 qual SINGL region"
 382..396
 /note"<30 qual SINGL region"
 412..416
 /note"<30 qual SINGL region"
 428..432
 /note"<30 qual SINGL region"
 487..575
 /note="single clone coverage"
 494..498
 /note"<30 qual single clone coverage"
 528..532
 /note"<30 qual single clone coverage"
 541..545
 /note"<30 qual single clone coverage"
 586..590
 /note"<30 qual SINGL region"
 1021..1530
 /rpt_family="MER34B"
 complement(1726..2118)
 /rpt_family="MER57A"
 complement(2570..2604)
 /rpt_family="L2"
 2770..3093
 /rpt_family="ALuY"
 3317..3612
 /rpt_family="ALuJo"
 complement(4209..4961)
 /rpt_family="L1MA4"
 6239..6415
 /rpt_family="MIR"
 7440..7484
 /rpt_family="(TCTCC)n"
 7570..7782
 /rpt_family="MIR"
 complement(7800..7896)
 /rpt_family="MIR"
 7944..8436
 /rpt_family="MLT1K"
 complement(8841..9063)
 /rpt_family="MIR"

unsure complement(8992..9086)
 /note="single clone coverage"
 unsure complement(8996..9001)
 /note"<30 qual single clone coverage"
 unsure complement(9035..9039)
 /note"<30 qual single clone coverage"
 unsure complement(9052..9063)
 /note"<30 qual single clone coverage"
 unsure complement(9120..9137)
 /note"<30 qual SINGL region"
 repeat_region complement(9179..9316)
 /rpt_family="FAM"
 9436..9567
 /rpt_family="MIR"
 9570..9603
 /rpt_family="(CAT)n"
 9729..9756
 /rpt_family="(CAAAA)n"
 repeat_region complement(10861..10957)
 /rpt_family="HAL1b"
 repeat_region complement(11110..11325)
 /rpt_family="LIME"
 repeat_region complement(11327..11603)
 /rpt_family="AluJo"
 11447..11452
 /note"<30 qual SINGL region"
 11457..11461
 /note"<30 qual SINGL region"
 repeat_region complement(11728..11816)
 /rpt_family="MIR3"
 repeat_region complement(11822..12123)
 /rpt_family="AluSx"
 12056..12061
 /note"<30 qual SINGL region"
 12129..12136
 /note"<30 qual SINGL region"
 12168..12174
 /note"<30 qual SINGL region"
 Alignment Scores:
 Pred. No.: 663 Length: 141899
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.55% Indels: 0
 DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AC087457 (1-141899)

QY 97 PheLeuArgGlyAsnProAsnLeuSer 105

Db 62431 TTCTTAAGGGCAACCCCAACCTGTCT 62457

RESULT 60

AC123206

LOCUS

DEFINITION

AC123206

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

150147 bp DNA linear HTG 17-JUL-2002
 Rattus norvegicus clone CH230-137M23, *** SEQUENCING IN PROGRESS
 *** 74 unordered pieces.

AC123206

AC123206.2 GI:21671690

HTG; HTGS_PHASE1.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 150147)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,

Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,

Barbata, J., Benton, J., Bimane, K., Blankenburg, K., Bonnin, D.,

Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,

Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,

Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,

```

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Hollaway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karissone,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
Lozador,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,X., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 150147)
Worley,K.C.
Direct Submission
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 150147)
Worley,K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 2, 2002 this sequence version replaced gi:21239988.
----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPGZ
Center clone name: CH230-137M23
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 85843 bases at least Q40
Consensus quality: 93252 bases at least Q30
Consensus quality: 97242 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 74 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

```

* as soon as it is available and the accession number will
* be preserved.
*
1 1454: contig of 1454 bp in length
* 1455: gap of unknown length
* 1555: contig of 1679 bp in length
* 3234: gap of unknown length
* 3334: contig of 1128 bp in length
* 4461: gap of unknown length
* 4561: gap of unknown length
* 5979: contig of 1418 bp in length
* 5980: gap of unknown length
* 6080: contig of 1602 bp in length
* 7682: gap of unknown length
* 7782: contig of 1610 bp in length
* 9392: gap of unknown length
* 9492: contig of 1233 bp in length
* 10725: gap of unknown length
* 10825: contig of 1177 bp in length
* 12002: gap of unknown length
* 12102: contig of 1460 bp in length
* 13561: gap of unknown length
* 13662: gap of unknown length
* 14708: contig of 1047 bp in length
* 14808: gap of unknown length
* 14809: gap of unknown length
* 14809: contig of 1072 bp in length
* 15881: gap of unknown length
* 15981: gap of unknown length
* 17340: contig of 1360 bp in length
* 17341: gap of unknown length
* 17441: gap of unknown length
* 19078: contig of 1638 bp in length
* 19079: gap of unknown length
* 19179: contig of 1691 bp in length
* 20869: gap of unknown length
* 20870: gap of unknown length
* 22724: contig of 1755 bp in length
* 22725: gap of unknown length
* 22824: gap of unknown length
* 23923: contig of 1099 bp in length
* 23924: gap of unknown length
* 24023: gap of unknown length
* 25106: contig of 1083 bp in length
* 25206: gap of unknown length
* 25207: contig of 1353 bp in length
* 26559: gap of unknown length
* 26560: gap of unknown length
* 27226: contig of 1067 bp in length
* 27227: gap of unknown length
* 27826: gap of unknown length
* 27827: contig of 1059 bp in length
* 28885: gap of unknown length
* 28886: contig of 1892 bp in length
* 30877: gap of unknown length
* 30878: gap of unknown length
* 32206: contig of 1229 bp in length
* 32207: gap of unknown length
* 33207: contig of 1589 bp in length
* 33895: gap of unknown length
* 33995: gap of unknown length
* 35594: contig of 1599 bp in length
* 35595: gap of unknown length
* 37941: contig of 2247 bp in length
* 37942: gap of unknown length
* 38041: contig of 1903 bp in length
* 38042: gap of unknown length
* 40044: gap of unknown length
* 41049: contig of 1005 bp in length
* 41149: gap of unknown length
* 41150: contig of 1426 bp in length
* 42575: gap of unknown length
* 42576: gap of unknown length
* 43884: contig of 1209 bp in length
* 43885: gap of unknown length
* 43985: contig of 1790 bp in length
* 45774: gap of unknown length
* 45775: gap of unknown length
* 45875: contig of 1554 bp in length
* 47429: gap of unknown length
* 47528: gap of unknown length
* 48199: contig of 1291 bp in length
* 48200: gap of unknown length
* 48919: gap of unknown length
* 50317: contig of 1398 bp in length
* 50417: gap of unknown length
* 50418: contig of 1145 bp in length
* 51562: gap of unknown length
* 51563: gap of unknown length
* 52781: contig of 1119 bp in length
* 52861: gap of unknown length
* 52882: contig of 1768 bp in length

```



```
repeat_region 8761..9067 /note="AluSq repeat: matches 1..306 of consensus"
repeat_region 9068..9174 /note="AluMb6 repeat: matches 5170..5305 of consensus"
repeat_region 9193..9350 /note="FAM repeat: matches 134..43 of consensus"
repeat_region 9353..9969 /note="FAM repeat: matches 5311..5919 of consensus"
repeat_region 9971..10354 /note="MLR1B repeat: matches 1..388 of consensus"
repeat_region 10467..10778 /note="AluMb repeat: matches 1..308 of consensus"
repeat_region 11491..11674 /note="L1MB1 repeat: matches 5955..6163 of consensus"
repeat_region 11708..11915 /note="L1MB8 repeat: matches 5938..6166 of consensus"
repeat_region 12438..12465 /note="L2 repeat: matches 2579..2604 of consensus"
repeat_region 12466..12764 /note="AluSc repeat: matches 1..300 of consensus"
repeat_region 12765..12905 /note="L2 repeat: matches 2604..2750 of consensus"
repeat_region 12921..13018 /note="FLAM_A repeat: matches 37..133 of consensus"
repeat_region 13172..13375 /note="MIR repeat: matches 53..262 of consensus"
repeat_region 14711..14894 /note="FRAM repeat: matches 6..174 of consensus"
repeat_region 15163..15460 /note="AluJo repeat: matches 1..300 of consensus"
repeat_region 15634..15940 /note="AluSx repeat: matches 2..305 of consensus"
repeat_region 16932..16959 /note="L4 copies 2 mer ca 92% conserved"
repeat_region 17439..17737 /note="AluSq repeat: matches 1..310 of consensus"
repeat_region 19260..19611 /note="L2 repeat: matches 2102..2471 of consensus"
repeat_region 20009..20171 /note="FRAM repeat: matches 9..173 of consensus"
repeat_region 20172..20223 /note="L3 copies 4 mer gggg 90% conserved"
repeat_region 20810..21105 /note="AluSq repeat: matches 1..300 of consensus"
repeat_region 21217..21351 /note="MIR repeat: matches 84..221 of consensus"
repeat_region 21442..21747 /note="AluTb repeat: matches 1..297 of consensus"
repeat_region 21893..22005 /note="MIR repeat: matches 96..212 of consensus"
repeat_region 22080..22198 /note="MER5B repeat: matches 1..126 of consensus"
repeat_region 22201..22232 /note="L6 copies 2 mer tg 96% conserved"
repeat_region 22203..22230 /note="L7 copies 4 mer ttgtg 100% conserved"
repeat_region 22233..22514 /note="AluY repeat: matches 1..282 of consensus"
repeat_region 22525..22575 /note="MER5B repeat: matches 125..175 of consensus"
repeat_region 24124..24384 /note="AluTb repeat: matches 5..302 of consensus"
repeat_region 24472..24620 /note="MLR1B repeat: matches 1..151 of consensus"
repeat_region 24987..25118 /note="MIR repeat: matches 1..137 of consensus"
repeat_region 25742..26046 /note="AluSq repeat: matches 1..305 of consensus"
repeat_region 27815..28096 /note="AluJo repeat: matches 1..282 of consensus"
repeat_region 28120..28430 /note="AluSx repeat: matches 1..311 of consensus"
repeat_region 28638..28718 /note="L2 repeat: matches 2663..2745 of consensus"
30584..30857 /note="AluJo repeat: matches 1..275 of consensus"
30892..31125 /note="MIR repeat: matches 12..255 of consensus"
31641..31763 /note="FLAM_A repeat: matches 1..122 of consensus"
31791..32009 /note="MIR repeat: matches 18..255 of consensus"
32551..32851 /note="AluSx repeat: matches 1..301 of consensus"
32866..33173 /note="AluSx repeat: matches 1..312 of consensus"
33511..33615 /note="U6 repeat: matches 3..107 of consensus"
33831..34127 /note="AluSx repeat: matches 1..295 of consensus"
34169..34263 /note="MIR repeat: matches 159..247 of consensus"
34222..34281 /note="L2 repeat: matches 2692..2747 of consensus"
34351..34836 /note="AluSx repeat: matches 2..297 of consensus"
34959..35034 /note="L2 repeat: matches 2630..2707 of consensus"
35209..35340 /note="MIR repeat: matches 112..254 of consensus"
35421..35662 /note="AluJo repeat: matches 57..300 of consensus"
35663..35692 /note="L5 copies 2 mer ga 86% conserved"
35772..36094 /note="AluSx repeat: matches 1..309 of consensus"
36095..36335 /note="L2 repeat: matches 2663..2745 of consensus"

Alignment Scores:
Pred. NO.: 713 Length: 154288
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.55% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x ALS13164 (1-154288)
QY 125 GlyLeuArgGluHisArgAlaGly 133
Db 18890 GGCTTAAGAGGTTACATAGGCTGGA 18916

RESULT 62
AC113662 158016 bp DNA linear HTG 13-JUL-2002
LOCUS Rattus norvegicus clone CH230-15401, *** SEQUENCING IN PROGRESS
DEFINITION *** 56 unordered pieces.
AC113662
VERSION AC113662.3 GI:21743470
KEYWORDS HTG: HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 158016)
AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., AmaralTunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.B., Dathorne,S.R., David,R.,
Devilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
```

```

Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegé,H.,
Lozadó,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 158016)
Worley,K.C.
Direct Submission
Submitted (05-MAR-2002) Human Genome Sequencing Center, Department
Of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 158016)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
Of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced gi:19525824.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRUR
Center clone name: CH230-15401
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 116208 bases at least Q40
Consensus quality: 119847 bases at least Q30
Consensus quality: 122552 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1728: contig of 1728 bp in length
* 1729 1828: gap of unknown length

```

```

1829
3041
3141
4238: contig of 1098 bp in length
4338: gap of unknown length
5474: gap of unknown length
5575: contig of 1360 bp in length
6935
7035
8042: contig of 1008 bp in length
8143
9327: contig of 1185 bp in length
9427: gap of unknown length
9428
10467: contig of 1040 bp in length
10468
11560: contig of 1093 bp in length
11661
13009: contig of 1249 bp in length
13010
13109: gap of unknown length
13110
14454: contig of 1345 bp in length
14455
15646: contig of 1092 bp in length
15647
16209: contig of 1074 bp in length
16210
16920: gap of unknown length
16921
18633: contig of 1713 bp in length
18634
18733: gap of unknown length
20290: contig of 1557 bp in length
20291
20390: gap of unknown length
20391
21752: contig of 1362 bp in length
21753
21852: gap of unknown length
21853
22933: contig of 1081 bp in length
22934
23033: gap of unknown length
23034
25044: contig of 2011 bp in length
25045
25144: gap of unknown length
25145
26595: contig of 1451 bp in length
26596
26995: gap of unknown length
28801: contig of 2106 bp in length
28901: gap of unknown length
28902
30135: contig of 1234 bp in length
30235: gap of unknown length
30236
32675: contig of 2440 bp in length
32676
34111: contig of 1336 bp in length
34112
34211: gap of unknown length
34212
36075: contig of 1864 bp in length
36076
36175: gap of unknown length
36176
38430: contig of 2255 bp in length
38431
38530: gap of unknown length
38531
40640: contig of 2110 bp in length
40641
43028: contig of 2288 bp in length
43029
43128: gap of unknown length
43129
45507: contig of 2379 bp in length
45508
45607: gap of unknown length
47252: contig of 1645 bp in length
47253
48961: contig of 1609 bp in length
48962
49061: gap of unknown length
49062
51776: contig of 2715 bp in length
51777
51876: gap of unknown length
53989: contig of 2113 bp in length
53990
54099: contig of 2410 bp in length
54099
56499: contig of 3068 bp in length
56500
59667: gap of unknown length
59668
59767: gap of unknown length
62514: contig of 2747 bp in length
62515
64470: gap of unknown length
64470
64570: contig of 1856 bp in length
64571
64571
66820: contig of 2250 bp in length
66821
68862: contig of 1942 bp in length
68862

```

```

* 68863 68962: gap of unknown length
* 68963 71828: contig of 2866 bp in length
* 71828 71928: gap of unknown length
* 71928 74748: contig of 2820 bp in length
* 74748 74848: gap of unknown length
* 74848 78059: contig of 3211 bp in length
* 78059 78159: gap of unknown length
* 78159 82256: contig of 4097 bp in length
* 82256 82357: gap of unknown length
* 82357 85575: contig of 3219 bp in length
* 85575 85675: gap of unknown length
* 85675 89332: contig of 3657 bp in length
* 89332 89432: gap of unknown length
* 89432 93853: contig of 4421 bp in length
* 93853 93953: gap of unknown length
* 93953 97732: contig of 3779 bp in length
* 97732 97832: gap of unknown length
* 97832 102363: contig of 4531 bp in length
* 102363 102463: gap of unknown length
* 102463 109116: contig of 6553 bp in length
* 109116 109117: gap of unknown length
* 109117 115198: contig of 6082 bp in length
* 115198 115298: gap of unknown length
* 115298 120793: contig of 5495 bp in length
* 120793 120893: gap of unknown length
* 120893 125338: contig of 4445 bp in length
* 125338 125439: gap of unknown length
* 125439 131505: contig of 6067 bp in length
* 131505 131506: gap of unknown length
* 131506 138366: contig of 6761 bp in length
* 138366 138467: gap of unknown length
* 138467 144803: contig of 6337 bp in length

```

Alignment Scores:

```

Pred. No.: 728 Length: 158016
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.55% Indels: 0
DB: 2 Gaps: 0

```

US-09-966-880A-8 (1-198) x AC113662 (1-158016)

Qy 175 GlnLeuArgArgIleLeuLeuProLeu 183

Db 151669 CAGCTACGTAGAAATTTACTCTTTG 151695

RESULT 63

AP002436/c

LOCUS Homo sapiens genomic DNA, chromosome 11q clone:RP11-627G1, complete sequences. PRI 22-FEB-2001

ACCESSION AP002436.3 GI:13094217

VERSION HTG.

KEYWORDS Homo sapiens DNA, clone:RP11-627G1.

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE 1 (sifts) Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Homo sapiens genomic DNA

JOURNAL Published Only in Database (2000)

REFERENCE 2 (bases 1 to 173509)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suihoro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)

COMMENT On Feb 21, 2001 this sequence version replaced gi:11071943.

FEATURES

```

Location/Qualifiers
source
1..173509
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-627G1"
BASE COUNT 53330 a 37471 c 36148 g 46560 t
ORIGIN

```

Alignment Scores:

```

Pred. No.: 789 Length: 173509
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.55% Indels: 0
DB: 9 Gaps: 0

```

US-09-966-880A-8 (1-198) x AP002436 (1-173509)

Qy 39 AlaThrSerPheSerLeuAspPhegly 47

Db 82024 GCACCACTTCTCTCTAGATTTTGGT 81998

RESULT 64

AC068190

LOCUS Homo sapiens

DEFINITION Homo sapiens chromosome 11 clone RP11-419E16 map 11, WORKING DRAFT

ACCESSION AC068190

VERSION AC068190.2 GI:8347999

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 176053)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 11, clone RP11-419E16

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 176053)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F., Bozulavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castile,A., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatus,A., Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehouckv,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trifilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (29-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 176053)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Bozulavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Farooq,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
 Iliev,I., Johnson,R., Jones,C., Karatas,A., Laroque,K.,
 Lanazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
 McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
 Severy,P., Sougnuez,C., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
 Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 8, 2000 this sequence version replaced gi:7670233.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9913

Center clone name: 419_E16

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 170544 bases at least Q40

Consensus quality: 173410 bases at least Q30

Consensus quality: 174372 bases at least Q20

Insert size: 176000; agarose-fp

Insert size: 175153; sum-of-contigs

Quality coverage: 5.0 in Q20 bases; agarose-fp

Quality coverage: 5.1 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1361: contig of 1361 bp in length
 * 1362 1461: gap of 100 bp
 * 1462 4481: contig of 3020 bp in length
 * 4482 4581: gap of 100 bp
 * 4582 12932: contig of 8351 bp in length
 * 12933 13032: gap of 100 bp
 * 13033 27303: contig of 14271 bp in length
 * 27304 27403: gap of 100 bp
 * 27404 39940: contig of 12537 bp in length
 * 39941 40040: gap of 100 bp
 * 40041 60017: contig of 19977 bp in length
 * 60018 60117: gap of 100 bp
 * 60118 83591: contig of 23474 bp in length
 * 83592 83691: gap of 100 bp
 * 83692 108521: contig of 24830 bp in length
 * 108522 108621: gap of 100 bp
 * 108622 139440: contig of 30819 bp in length
 * 139441 139540: gap of 100 bp
 * 139541 176053: contig of 36513 bp in length.
 * Location/Qualifiers
 * 1. 176053.

/organism="Homo sapiens"

/db_xref="taxon:9606"

FEATURES

source

/chromosome="11"
 /map="11"
 /clones="RP11-419E16"
 /clone_lib="RPCI-11 Human Male BAC"
 1. 1361
 /note="assembly_fragment"
 1462. 4481
 /note="assembly_fragment"
 4582. 12932
 /note="assembly_fragment"
 13033. 27303
 /note="assembly_fragment"
 clone_end:T7
 vector_side:left
 27404. 39940
 /note="assembly_fragment"
 40041. 60017
 /note="assembly_fragment"
 60118. 83591
 /note="assembly_fragment"
 83692. 108521
 /note="assembly_fragment"
 108622. 139440
 /note="assembly_fragment"
 139541. 176053
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 49213 a 37507 c 38347 g 50081 t 905 others
 BASE COUNT
 ORIGIN

Alignment Scores:

Pred. No.: 799 Length: 176053
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.53% Indels: 0
 DB: 2 Gaps: 0

US-09-966-880A-8 (1-198) x AC068190 (1-176053)

Qy 39 AlaThrSerPheSerLeuAspPheGly 47

Db 117796 GCAACACAGTTCTCTAGATTTTGGT 117822

RESULT 65

AC127421

LOCUS AC127421 179510 bp DNA linear HTG 17-JUL-2002

DEFINITION Mus musculus chromosome UNK clone RP24-314F1, WORKING DRAFT

SEQUENCE, 8 unordered pieces.

AC127421

VERSION AC127421.1 GI:21759528

KEYWORDS HTG; HTGS_PHRASE1; HTGS_DRAFT.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 179510)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

AUTHORS McPherson,J.D. and Waterston,R.H.

TITLE The sequence of Mus musculus clone

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 179510)

AUTHORS McPherson,J.D. and Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (15-JUL-2002) Genome Sequencing Center, 4444 Forest Park

REFERENCE 3 (bases 1 to 179510)

AUTHORS McPherson,J.D. and Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (17-JUL-2002) Genome Sequencing Center, 4444 Forest Park

Parkway, St. Louis, MO 63108, USA

COMMENT

----- Genome Center -----

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BB0314F01

----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Assembly: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 181162 bases at least Q40
Consensus quality: 181808 bases at least Q30
Consensus quality: 182185 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1378: contig of 1378 bp in length
1379 1478: gap of unknown length
1479 4229: contig of 2751 bp in length
4230 4329: gap of unknown length
4330 10616: contig of 6287 bp in length
10617 10716: gap of unknown length
10717 34165: contig of 23449 bp in length
34166 34265: gap of unknown length
34266 55394: contig of 21129 bp in length
55395 5494: gap of unknown length
5495 92891: contig of 37397 bp in length
92892 129685: contig of 36694 bp in length
129686 129785: gap of unknown length
129786 179510: contig of 4725 bp in length.

FEATURES
source

1..179510
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clones="RP24-314F1"
1..1378
/note="assembly_name:Contig17"
1479..4229
/note="assembly_name:Contig18"
4330..10616
/note="assembly_name:Contig20"
10717..34165
/note="assembly_name:Contig21"
34266..55394
/note="assembly_name:Contig22"
55495..92891
/note="assembly_name:Contig23"
92992..129685
/note="assembly_name:Contig24"
129786..179510
/note="assembly_name:Contig25"
BASE COUNT 51663 a 37702 c 36620 g 52806 t 719 others
ORIGIN

Alignment Scores:
Pred. No.: 812 Length: 179510
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.55% Indels: 0
DB: 2 Gaps: 0

US-09-966-880A-8 (1-198) x AC127421 (1-179510)
Qy 103 AsnLeuSerLeuArgIlePheThrAla 111
|||||
Db 63336 AACCTCTCTCTCGGATATTACTGCT 63362

RESULT 66

AC020857
LOCUS AC020857 180568 bp DNA linear HTG 16-FEB-2000
DEFINITION Mus musculus clone RP23-163F21, LOW-PASS SEQUENCE SAMPLING.
AC020857
VERSION AC020857.2 GI:6980198
KEYWORDS HTG; HTGS PHASE0.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 180568)
DOE Joint Genome Institute.
Unpublished
2 (bases 1 to 180568)
DOE Joint Genome Institute.
Direct Submission
Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Feb 16, 2000 this sequence version replaced gi:6686451.
* NOTE: This record contains 205 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 796: contig of 796 bp in length
gap of unknown length
797 1094: contig of 298 bp in length
gap of unknown length
1095 2104: contig of 1010 bp in length
gap of unknown length
2105 3041: contig of 937 bp in length
gap of unknown length
3042 3673: contig of 632 bp in length
gap of unknown length
3674 4251: contig of 578 bp in length
gap of unknown length
4252 4677: contig of 426 bp in length
gap of unknown length
4678 5588: contig of 911 bp in length
gap of unknown length
5589 6457: contig of 869 bp in length
gap of unknown length
6458 6775: contig of 318 bp in length
gap of unknown length
6776 7759: contig of 984 bp in length
gap of unknown length
7760 8466: contig of 707 bp in length
gap of unknown length
8467 8733: contig of 267 bp in length
gap of unknown length
8734 9823: contig of 1090 bp in length
gap of unknown length
9824 10617: contig of 794 bp in length
gap of unknown length
10618 11286: contig of 669 bp in length
gap of unknown length
11287 11422: contig of 136 bp in length
gap of unknown length
11423 11683: contig of 261 bp in length
gap of unknown length


```

Qy 103 AsnLeuSerLeuArgIlePheThrAla 111
|||||
Db 146416 AACCTCTCTCTCGGATATTTACTGCT 146442

RESULT 67
AP003402c 187948 bp DNA linear PRI 06-NOV-2001
LOCUS Homo sapiens genomic DNA, chromosome 11q clone:RP11-661121,
DEFINITION complete sequence.
ACCESSION AP003402
VERSION AP003402.2 GI:16751506
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-661121.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (2001)
REFERENCE
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2001) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@psc.riken.go.jp, URL:http://hgsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Nov 5, 2001 this sequence version replaced gi:13366105.
FEATURES
source
1..187948
Location/Qualifiers
1..187948
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11q"
/clone="RP11-661121"
BASE COUNT 55717 a 41857 c 40668 g 49706 t
ORIGIN
|||||
Alignment Scores:
Pred. No.: 845 Length: 187948
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.55% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AP003402 (1-187948)
Qy 39 AlatrSerPheSerLeuAspPheGly 47
|||||
Db 159621 GCACACAGTTCTCTCTAGATTTTGGT 159595

RESULT 68
AC007083 188067 bp DNA linear INV 17-MAR-2001
LOCUS Drosophila melanogaster, chromosome 2L, region 33A-, BAC clone
DEFINITION BACR43K24, complete sequence.
ACCESSION AC007083
VERSION AC007083.11 GI:13374638
KEYWORDS HTG.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,

```

```

Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farlan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Shaplanon,M., Strong,R., Svirska,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 2L, region 33A-
Unpublished
2 (bases 1 to 188067)
Celniker,S.E., Aghayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Madxa,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirska,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zierun,L.L. and
Rubin,G.M.
Direct Submission
Submitted (16-MAR-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Mar 17, 2001 this sequence version replaced gi:7264773.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgpf@fruitfly.berkeley.edu.
FEATURES
source
1..188067
Location/Qualifiers
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="33A-"
/clone="BACR43K24 (D575)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
pBACE3.6)"
BASE COUNT 53731 a 40263 c 40250 g 53823 t
ORIGIN
|||||
Alignment Scores:
Pred. No.: 846 Length: 188067
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.55% Indels: 0
DB: 3 Gaps: 0

US-09-966-880A-8 (1-198) x AC007083 (1-188067)
Qy 117 GluAspArgLysAlaGluProGluGly 125
|||||
Db 105148 GAGGACAGAAAGGAGAGAGCCAGAGGTT 105174

RESULT 69
AL840638 188136 bp DNA linear HTG 17-JUL-2002
LOCUS Danio rerio clone DKEY-224J12, *** SEQUENCING IN PROGRESS ***
DEFINITION unordered pieces.
ACCESSION AL840638
VERSION AL840638.3 GI:21911823
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Danio rerio.
ORGANISM Danio rerio

```


Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 188136)
 Burton, J.
 Direct Submission
 Submitted (14-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
 Clone requests: clonerequest@sanger.ac.uk
 On Jul 19, 2002 this sequence version replaced gi:21748404.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zface@sanger.ac.uk
 ----- Project Information
 Center project name: zk224J12
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 181961 bases at least Q40
 Consensus quality: 184055 bases at least Q30
 Consensus quality: 185651 bases at least Q20
 Insert size: 187236; sum-of-contigs
 Insert size: 157949; 24.5% error; agarose-fp
 Quality coverage: 12.05x in Q20 bases; sum-of-contigs Quality
 coverage: 14.71x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 4964: contig of 4964 bp in length
 * 4965 5064: gap of 100 bp
 * 5065 13025: contig of 7961 bp in length
 * 13026 13125: gap of 100 bp
 * 13126 64622: contig of 51497 bp in length
 * 64623 64722: gap of 100 bp
 * 64723 73250: contig of 8528 bp in length
 * 73251 125837: contig of 52487 bp in length
 * 125838 125937: gap of 100 bp
 * 125938 156004: contig of 30067 bp in length
 * 156005 156104: gap of 100 bp
 * 156105 159322: contig of 3218 bp in length
 * 159323 159422: gap of 100 bp
 * 159423 161843: contig of 2421 bp in length
 * 161844 161943: gap of 100 bp
 * 161944 165041: contig of 3098 bp in length
 * 165042 165141: gap of 100 bp
 * 165142 188136: contig of 22995 bp in length.
 Location/Qualifiers
 1..188136
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone="DKEX-224J12"
 /clone_lib="Daniokeypilot"
 1..4964
 /note="assembly_fragment:00064
 fragment_chain:1"
 5065..13025
 /note="assembly_fragment:02032
 fragment_chain:1"
 13126..64622
 /note="assembly_fragment:00273
 fragment_chain:1"
 64723..73250
 /note="assembly_fragment:02919
 fragment_chain:1"

misc_feature 73351..125837
 /note="assembly_fragment:01790
 fragment_chain:1"
 125838..156004
 /note="assembly_fragment:00407"
 156105..159322
 /note="assembly_fragment:00768"
 159423..161843
 /note="assembly_fragment:02758"
 161944..165041
 /note="assembly_fragment:02989"
 165142..188136
 /note="assembly_fragment:03788..0"
 BASE COUNT 60355 a 33818 c 33700 g 59356 t 907 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 846 Length: 188136
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.55% Indels: 0
 DB: 2 Gaps: 0
 US-09-966-880A-8 (1-198) x AL840638 (1-188136)
 Qy 59 LeuLeuPheLeuArgTyrIleSerAsp 67
 |||||
 Db 35708 CTGCTTTTCCTCCGCTACATCTCAGAT 35734
 |||||
 RESULT 70
 AC036188
 LOCUS
 DEFINITION Homo sapiens chromosome 11 clone RP11-287H12 map 11, WORKING DRAFT
 SEQUENCE, 19 unordered pieces.
 AC036188 190960 bp DNA linear HTG 24-AUG-2002
 AC036188.2 GI:7770585
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 190960)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Unpublished
 JOURNAL
 2 (bases 1 to 190960)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bida, F.,
 Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrim, J., Meneus, D., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neil, D., Olivari, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triglio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zainou, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 190960)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

```

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Navlor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 12, 2000 this sequence version replaced gi:7523857.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9252
Center clone name: 287_H12
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 178632 bases at least Q40
Consensus quality: 185374 bases at least Q30
Consensus quality: 187834 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 189160; sum-of-contigs
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1629: contig of 1629 bp in length
* 1630 1729: gap of 100 bp
* 1730 3289: contig of 1560 bp in length
* 3290 3389: gap of 100 bp
* 3390 4902: contig of 1513 bp in length
* 4903 5002: gap of 100 bp
* 5003 8987: contig of 3985 bp in length
* 8988 9087: gap of 100 bp
* 9088 12144: contig of 3057 bp in length
* 12145 12244: gap of 100 bp
* 12245 16220: contig of 3976 bp in length
* 16221 16320: gap of 100 bp
* 16321 20626: contig of 4306 bp in length
* 20627 20726: gap of 100 bp
* 20727 25379: contig of 4653 bp in length
* 25380 25479: gap of 100 bp
* 25480 32157: contig of 6678 bp in length
* 32158 32257: gap of 100 bp
*
* 32258 42184: contig of 9927 bp in length
* 42185 42284: gap of 100 bp
* 42285 48095: contig of 5811 bp in length
* 48096 48195: gap of 100 bp
* 48196 59553: contig of 11358 bp in length
* 59554 59653: gap of 100 bp
* 59654 70846: contig of 11193 bp in length
* 70847 70946: gap of 100 bp
* 70947 87757: contig of 18811 bp in length
* 87758 87857: gap of 100 bp
* 87858 103500: contig of 15643 bp in length
* 103501 103600: gap of 100 bp
* 103601 119609: contig of 16009 bp in length
* 119610 119709: gap of 100 bp
* 119710 138055: contig of 18346 bp in length
* 138056 138155: gap of 100 bp
* 138156 158360: contig of 20205 bp in length
* 158361 158460: gap of 100 bp
* 158461 190960: contig of 32500 bp in length.
*
FEATURES
Location/Qualifiers
1..190960
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone_lib="RPC1-11 Human Male BAC"
1..1629
/note="assembly_fragment"
1730..3289
/note="assembly_fragment"
3390..4902
/note="assembly_fragment"
5003..8987
/note="assembly_fragment"
9088..12144
/note="assembly_fragment"
clone_end:T7
vector_side:left
12245..16220
/note="assembly_fragment"
16321..20626
/note="assembly_fragment"
20727..25379
/note="assembly_fragment"
25480..32157
/note="assembly_fragment"
32258..42184
/note="assembly_fragment"
42285..48095
/note="assembly_fragment"
clone_end:SP6
vector_side:right
48196..59553
/note="assembly_fragment"
59654..70846
/note="assembly_fragment"
70947..87757
/note="assembly_fragment"
87858..103500
/note="assembly_fragment"
103601..119609
/note="assembly_fragment"
119710..138055
/note="assembly_fragment"
138156..158360
/note="assembly_fragment"
158461..190960
/note="assembly_fragment"
BASE COUNT 54008 a 39431 c 41327 g 54391 t 1803 others
ORIGIN
Alignment Scores:

```



```

/number=6
13243..15390
/genes="B14D6.020"
CDS
13243..15390
/genes="B14D6.020"
/notes="similarity to product:unknown, clone pp2464, homo sapiens, TREMBL:AF218002.1"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAD21209.1"
/db_xref="GI:18375979"
/translation="MSLHVPPSAQKNVLDLDDYDNFDTADDVPISTSKTPARKEQ
GLSTVDLGDSEQLKSQARTMTQKTPELDIDSGDFLIDITRSTIRGERTS
EPVQRQLSPVSESNPSKRLAQTHTPAISQSLARLGPSSRSRVTGESSSTPD
ISGPNISAAQPTTYHNDNDNDPPASPPPRETREQKGKGPASTTDPVINIDDDDDP
FADSLPKVNPGRCPRAATTAANDPISSSAPLPKPSPHRSLGRSSRSPLESDS
EDIGADPLSNDEDDFPDLSLFPANRLGNDLTGTPFTKASATTTGATKTPAKG
GPKTAERAKEREDKALEKSAERKQERERQKAEQKQAAEAANKLRK
ETATPEMDVLPSSLPPTKIQMEFLKKIDYKEINTWTSPDNNVRRRAVKSRYNE
ERHVDPIPETIEKIIIVLPAAEFKAKLMAEGHNLEAHVLKQVRFNHOIYL
TEGLKLLSSNNRNNDPASVRSRLAEDESASTSSSRRTNKKNDPMTTISEQIDA
ALLRLQLYSNQIOETTCLODPAHLQLFTQNVAVAPYKRQEDLYLMSAGFCMDSGQ
VRTALGTAEAYRMLQEVARTIAPAMGIANYIPRVQQLVRALEEGGPGTLEDIRKVI
NKEREVGKRVKAVSKRLWKIFTGRDEMSTEV"
13243..15390
/genes="B14D6.020"
/number=1
/genes="B14D6.020"
/genes="B14D6.030"
Join(17216..17318,17488..18161,18247..18445,18503..18567)
/genes="B14D6.030"
/notes="contains EST gb:BF072791"
/codon_start=1
/product="putative protein"
/protein_id="CAD21210.1"
/db_xref="GI:18375980"
/translation="MENLRPLLSLDIDRTIGSFDFLEFAENNESLMRLVFRHQFOR
TDMQRDSFGSAYSHCYEDKIELFLTYAFSSNETTALFLNAGFGYMRKDKMCKFDE
ISKVTCKVGHPLTLPYIRINVRRLNQRLHLEAAWQOMELFLACNTDEVNNIRKI
YNTSDSHQRLGRIGHADNRSMISVLSSAAALCEAIDELREMAWFSWLSDEKS
SEMTAHSQALKEERAQDLSIAGFTROLEORMHALESHTHMALESQNSQLNM
ATAQOQIDLSITTTTRDAKTMGLAFMGALFLPGTFMSAIFSNPFEEKIDKLWTFAA
AVPLTLITITFG"
17216..17318
/genes="B14D6.030"
/number=1
/genes="B14D6.030"
/genes="B14D6.030"
/number=1
/notes="35 bp tgtgtgtgtgtgtgtgtgtgt tandem repeat"
17409..17443
/genes="B14D6.030"
/number=2
18162..18246
/genes="B14D6.030"
/number=2
18247..18445
/genes="B14D6.030"
/number=3
18446..18502
/genes="B14D6.030"
/number=3
18503..18567
/genes="B14D6.030"
/number=4
19258..19377
/product="5S ribosomal RNA"
20697..21695

```

Alignment Scores:

Pred. No.: 890 Length: 19936
Score: 9.00 Matches: 9

```

Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.55% Indels: 0
DB: 8 Gaps: 0

US-09-966-880A-8 (1-198) x NCB14D6 (1-199386)

Qy 174 ArgGlnLeuArgArgileLeuLeuPro 182
Db 171653 AGACAACCTGGCGGCATCTTCTGCGCA 171679

RESULT 72
AC103892/c
LOCUS
DEFINITION
AC103892
AC103892.2 GI:20429784
HTG: HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 207584)
Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Lalic,P.,
Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Mastello,C., Maskeri,B., Mastrian,S.D.,
McCluskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
Thomas,P.J., Touchman,J.W., Tsurgan,C., Vogt,J.L., Walker,M.A.,
Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 207584)
Direct Submission
Submitted (30-NOV-2001) NIH Intramural Sequencing Center, 8717
Grovmont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 207584)
Direct Submission
Submitted (03-MAY-2002) NIH Intramural Sequencing Center, 8717
Grovmont Circle, Gaithersburg, MD 20877, USA
On May 3, 2002 this sequence version replaced gi:17155044.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: ckc
Center clone name: 108E11

```

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319

Consensus quality: 205765 bases at least Q40
 Consensus quality: 206455 bases at least Q30
 Consensus quality: 206619 bases at least Q20
 Insert size: 199000; agarose-fp
 Quality coverage: 10.48x in Q20 bases; agarose-fp
 Quality coverage: 10.09x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submittor.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

1 9523: contig of 9523 bp in length
 * 9524 9623: gap of unknown length
 * 9624 19034: contig of 9411 bp in length
 * 19035 19134: gap of unknown length
 * 19135 28903: contig of 9769 bp in length
 * 28904 29003: gap of unknown length
 * 29004 54011: contig of 25008 bp in length
 * 54012 54111: gap of unknown length
 * 54112 69522: contig of 15411 bp in length
 * 69523 69623: gap of unknown length
 * 69624 135448: contig of 65826 bp in length
 * 135449 135548: gap of unknown length
 * 135549 144120: contig of 8572 bp in length
 * 144121 144220: gap of unknown length
 * 144221 171771: contig of 27551 bp in length
 * 171772 171871: gap of unknown length
 * 171872 199518: contig of 27647 bp in length
 * 199519 199618: gap of unknown length
 * 199619 207584: contig of 7966 bp in length.

FEATURES

source

1. .207584
 /organism="Rattus norvegicus"
 /strain="Brown Norway"
 /db_xref="taxon:10116"
 /clone="RP31-108E11"
 /clone_lib="RP31"
 1. .41024
 /note="clone overlaps with GenBank Accession Number AC097001 clone RP31-358N14 (center project name ckb)"
 misc_feature
 1. .9523
 /note="assembly_fragment
 clone_end:17
 vector_side:left"
 misc_feature
 9624. .19034
 /note="assembly_fragment"
 misc_feature
 19135. .28903
 /note="assembly_fragment"
 misc_feature
 29004. .54011
 /note="assembly_fragment"
 misc_feature
 54112. .69522
 /note="assembly_fragment"
 misc_feature
 69623. .135448
 /note="assembly_fragment"
 misc_feature
 135549. .144120
 /note="assembly_fragment"
 misc_feature
 144221. .171771
 /note="assembly_fragment"
 misc_feature
 171872. .199518
 /note="assembly_fragment"
 misc_feature
 199619. .207584
 /note="assembly_fragment
 clone_end:SP6
 vector_side:right"
 BASE COUNT 60061 a 46035 c 45902 g 54686 t 900 others
 ORIGIN

Alignment Scores:

Pred. No.: 921 Length: 207584
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.55% Indels: 0
 DB: 2 Gaps: 0

US-09-966-880A-8 (1-198) x AC103892 (1-207584)

Qy 42 PheSerLeuAspPheGlyTyrLeuArg 50
 |||
 Db 181210 TTTTCTTTGGATTTGGTTATTACGG 181184

RESULT 73

AC091322

LOCUS

DEFINITION Mus musculus clone RP23-60D3, WORKING DRAFT SEQUENCE, 5 unordered pieces.

AC091322

AC091322.2 GI:14336533

VERSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS Mus musculus.

SOURCE Mus musculus

ORGANISM

REFERENCE

1 (bases 1 to 232951)

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

1 (bases 1 to 232951)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: LI3252

Center clone name: 60_D3

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 231176 bases at least Q40

Consensus quality: 231965 bases at least Q30

Consensus quality: 232319 bases at least Q20
 Insert size: 210000; agarose-fp
 Insert size: 232551; sum-of-contigs
 Quality coverage: 10.7 in Q20 bases; agarose-fp
 Quality coverage: 9.6 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 53: contig of 53 bp in length
 54 153: gap of 100 bp
 154 2590: contig of 2437 bp in length
 2591 2690: gap of 100 bp
 2691 11774: contig of 9084 bp in length
 11775 11874: gap of 100 bp
 11875 100524: contig of 88650 bp in length
 100525 100624: gap of 100 bp
 100625 232951: contig of 132327 bp in length.

FEATURES

source
 1. .232951
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-60D3"
 /clone_lib="RPC1-23 Female Mouse BAC"
 misc_feature
 1. .53
 /note="assembly_fragment"
 clone_end:sp6
 vector_side:left
 misc_feature
 154. .2590
 /note="assembly_fragment"
 2691. .11774
 /note="assembly_fragment"
 11875. .100524
 /note="assembly_fragment"
 misc_feature
 100625. .232951
 /note="assembly_fragment"
 clone_end:t7
 vector_side:right"
 BASE COUNT 70795 a 45474 c 46842 g 69434 t 406 others
 ORIGIN

Alignment Scores:
 Pred. No.: 1.02e+03 Length: 232951
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.55% Indels: 0
 DB: 2 Gaps: 0

US-09-966-880A-8 (1-198) x AC091322 (1-232951)

Qy 20 TrpAlaLySGlyArgArgGluThrTy 28

|||||

Db 143785 TGGCAAGGCGAGGAGGAGACTTAT 143811

RESULT 74
 LOCUS AE003632 244237 bp DNA linear INV 04-OCT-2000
 DEFINITION Drosophila melanogaster genomic scaffold 142000013386055 section 25
 of 63, complete sequence.

ACCESSION AE003632 AE002690

VERSION AE003632.2 GI:10728720

KEYWORDS HTG.

SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

1 (bases 1 to 244237)
 Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
 Ananides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
 George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
 Sutton,G.G., Wortman,J.R., Blazell,M.D., Zhang,Q., Chen,L.X.,
 Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
 Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
 Miklos,G.L., Abril,J.F., Aghayani,A., An,H.J.,
 Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
 Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
 Benos,P.V., Bereman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
 Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Hurlis,K.C.,
 Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
 Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
 Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
 Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., DunKov,B.C.,
 Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrier,S.,
 Fleischmann,W., Fowler,C., Gabriellian,A.E., Garg,N.S.,
 Gelbart,W.M., Glasser,K., Glodek,A., Gong,P., Gorrell,J.H., Gu,Z.,
 Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
 Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
 Wei,M.H., Ilegwam,C., Jaitani,M., Kalush,F., Karpen,G.H., Ke,Z.,
 Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
 Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
 Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattel,B., McIntosh,T.C.,
 McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
 Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,H., Murphy,L.,
 Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
 Nusskern,D.R., Pacieb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
 Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
 Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
 Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
 Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
 Weinstein,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
 Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
 Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
 Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
 The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

gen

mRNA

CDS

source

gen

mRNA

CDS

source

gen

mRNA

CDS

source

gen

mRNA

CDS

source

gen

mRNA

CDS

source

gen

mRNA

CDS

AUTHORS Xiang,Q.Y., Soltis,D.E. and Soltis,P.S.
 TITLE The eastern Asian and eastern and western North American floristic
 disjunction: congruent phylogenetic patterns in seven diverse
 genera

JOURNAL Mol. Phylogenet. Evol. 10 (2), 178-190 (1998)
 MEDLINE 99097283
 PUBMED 9878229

REFERENCE 2 (bases 1 to 245)
 AUTHORS Xiang,Q.-Y., Soltis,D.E. and Soltis,P.S.
 TITLE Direct Submission
 JOURNAL Submitted (02-APR-1997) Plant Biology, Ohio State University, 1735
 Neil Avenue, Columbus, OH 43210, USA

FEATURES
 source
 1..245
 Location/Qualifiers
 /organism="Trautvetteria carolinensis"
 /db_xref="taxon:46993"
 misc_RNA
 1..245
 /product="internal transcribed spacer 1"
 /note="ITS1"

BASE COUNT 65 a 65 c 64 g 50 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 31.3 Length: 245
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.04% Indels: 0
 DB: 0 Gaps: 0

US-09-966-880A-8 (1-198) x TCARITS1 (1-245)

Qy 174 ArgGlnLeuArgArgIleLeuLeu 181
 Db 173 GCCCAATTGCCCGGATTTTGTG 150

RESULT 76
 HSA279528
 LOCUS HSA279528 354 bp mRNA linear PRI 01-JUN-2000
 DEFINITION Homo sapiens mRNA for immunoglobulin mu heavy chain variable
 region, partial, clone CD27high-9.

ACCESSION AJ279528
 VERSION AJ279528.1 GI:6723553
 KEYWORDS IgM; IgM heavy chain; immunoglobulin mu heavy chain; variable
 region.

SOURCE
 human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 354)
 AUTHORS Dono,M., Zupo,S., Leanza,N., Mellioli,G., Fogli,M., Melagrana,A.,
 Chiorazzi,N. and Ferrarini,M.
 TITLE Heterogeneity of tonsillar subepithelial B lymphocytes, the splenic
 marginal zone equivalents
 JOURNAL J. Immunol. 164 (11), 5596-5604 (2000)
 MEDLINE 20281644
 PUBMED 10820234

REFERENCE 2 (bases 1 to 354)
 AUTHORS Dono,M.
 TITLE Direct Submission
 JOURNAL Submitted (11-JAN-2000) Dono M., Istituto Nazionale per la Ricerca
 sul Cancro, Immunologia Clinica, L.go Rosanna Benzi 10- Genova,
 ITALY

FEATURES
 source
 1..354
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone="CD27high-9"
 /cell_type="IgD-low SE B-lymphocyte"
 /tissue_type="tonsil"
 /rearranged
 1..354

gene

CDS
 /gene="IGHV4-34"
 <1..>354
 /gene="IGHV4-34"
 /codon_start=1
 /product="Immunoglobulin mu heavy chain variable region"
 /protein_id="CAB66358.1"
 /db_xref="GI:6723554"
 /translation="QVLLQMGAGLLKPKSETLSLTCAVYGGSFNDYYWSMIRQPHGKG
 LEWIGNLYYSGSTYNPGLKSRVTIFVDTSKNQFSLKLSSTAAADTAVYYCARHVA
 RRFYWGQGTLLTVSS"
 <1..>354
 /gene="IGHV4-34"
 /product="Immunoglobulin mu heavy chain variable region"

V_region
 74 a 99 c 98 g 83 t

BASE COUNT
 ORIGIN

Alignment Scores:
 Pred. No.: 43.1 Length: 354
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.04% Indels: 0
 DB: 0 Gaps: 0

US-09-966-880A-8 (1-198) x HSA279528 (1-354)

Qy 90 CysAlaArgHisValAlaAspPhe 97
 Db 283 TGTGGGAGACACGTCGCGGACTTT 306

RESULT 77
 AX389702/C
 LOCUS AX389702 500 bp DNA linear PAT 19-MAR-2002
 DEFINITION Sequence 4630 from Patent WO214500.
 ACCESSION AX389702
 VERSION AX389702.1 GI:19582832
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
 AUTHORS Escobedo,J., Garcia,P.D., Sudduth-Klinger,J., Reinhard,C.,
 Randazzo,F., Lamson,G., Scott,E.M., Zhang,G., Kassam,A., Pot,D. and
 Labat,I.
 TITLE Human genes and gene expression products
 JOURNAL Patent: WO 0214500-A 4630 21-FEB-2002;
 CHIRON CORPORATION (US) ; Hyseq Inc. (US)
 FEATURES Location/Qualifiers
 source
 1..500
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 227 a 75 c 72 g 126 t
 ORIGIN

Alignment Scores:
 Pred. No.: 58.1 Length: 500
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.04% Indels: 0
 DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x AX389702 (1-500)

Qy 59 LeuLeuPheLeuArgTyrIleSer 66
 Db 264 TTGTTGTTTCTAAGATACATATCT 241

RESULT 78
 HSC07H12
 LOCUS HSC07H12 569 bp DNA linear STS 01-MAR-2000
 DEFINITION H.sapiens STS from flow-sorted chromosome X random shear fragment,


```

sequence tagged site.
ACCESSION AL158650
VERSION AL158650.1 GI:7161252
KEYWORDS STS.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 569)
AUTHORS Hunt,S., Pavitt,B., Willey,D., Carter,N. and Ross,M.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-2000) E-mail contact: humquerry@sanger.ac.uk
COMMENT Marker STFX1703 (Primer A : CTCCTCAGAAGCATTCCTG; Primer B :
TGGTTTATGCTTGTGTTTCCCA; amplicon size : 184 bp) was developed from
a single pass sequencing read from H.sapiens flow-sorted chromosome
X random shear fragment, SCXpc07h12. Vector : pUC18 Site : SmaI
Further information : http://www.sanger.ac.uk/HGP/ChrX/.
FEATURES
Location/Qualifiers
1..569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="SCXpc07h12"
/sex="male"
/tissue_type="Lymphoblastoid cell line"
/clone_lib="SCXpc"
BASE COUNT 203 a 92 c 76 g 198 t
ORIGIN
Alignment Scores:
Pred. No.: 64.9 Length: 569
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 11 Gaps: 0

US-09-966-880A-8 (1-198) x HSC07H12 (1-569)
Qy 138 lleMetThrPheLysAspTyrPhe 145
|||||
Db 306 ATCATGACATTAAAGATTATTTT 329

RESULT 79
PM3C3B PM3C3B 653 bp DNA linear STS 09-MAR-2002
LOCUS Penicillium marneffei STS, clone pm3c3.b, sequence tagged site.
ACCESSION AL685634
VERSION AL685634.1 GI:19336682
KEYWORDS STS.
SOURCE Penicillium marneffei.
ORGANISM Penicillium marneffei
REFERENCE 1
AUTHORS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
Yuen,K.Y., Pascal,G., Wong,S., Glaser,P., Woo,P., Kunst,P.,
Cheung,E., Medigue,C. and Danchin,A.
TITLE Exploring the Penicillium marneffei genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 653)
AUTHORS Danchin,A. and Pascal,G.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
FEATURES
Location/Qualifiers
1..653
/organism="Penicillium marneffei"
/db_xref="taxon:37727"
/clone="pm3c3.b"
BASE COUNT 176 a 159 c 140 g 178 t
ORIGIN
Alignment Scores:
sequence tagged site.
ACCESSION AL158650
VERSION AL158650.1 GI:7161252
KEYWORDS STS.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 569)
AUTHORS Hunt,S., Pavitt,B., Willey,D., Carter,N. and Ross,M.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-2000) E-mail contact: humquerry@sanger.ac.uk
COMMENT Marker STFX1703 (Primer A : CTCCTCAGAAGCATTCCTG; Primer B :
TGGTTTATGCTTGTGTTTCCCA; amplicon size : 184 bp) was developed from
a single pass sequencing read from H.sapiens flow-sorted chromosome
X random shear fragment, SCXpc07h12. Vector : pUC18 Site : SmaI
Further information : http://www.sanger.ac.uk/HGP/ChrX/.
FEATURES
Location/Qualifiers
1..569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="SCXpc07h12"
/sex="male"
/tissue_type="Lymphoblastoid cell line"
/clone_lib="SCXpc"
BASE COUNT 203 a 92 c 76 g 198 t
ORIGIN
Alignment Scores:
Pred. No.: 64.9 Length: 569
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 11 Gaps: 0

US-09-966-880A-8 (1-198) x HSC07H12 (1-569)
Qy 138 lleMetThrPheLysAspTyrPhe 145
|||||
Db 306 ATCATGACATTAAAGATTATTTT 329

RESULT 79
PM3C3B PM3C3B 653 bp DNA linear STS 09-MAR-2002
LOCUS Penicillium marneffei STS, clone pm3c3.b, sequence tagged site.
ACCESSION AL685634
VERSION AL685634.1 GI:19336682
KEYWORDS STS.
SOURCE Penicillium marneffei.
ORGANISM Penicillium marneffei
REFERENCE 1
AUTHORS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
Yuen,K.Y., Pascal,G., Wong,S., Glaser,P., Woo,P., Kunst,P.,
Cheung,E., Medigue,C. and Danchin,A.
TITLE Exploring the Penicillium marneffei genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 653)
AUTHORS Danchin,A. and Pascal,G.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
FEATURES
Location/Qualifiers
1..653
/organism="Penicillium marneffei"
/db_xref="taxon:37727"
/clone="pm3c3.b"
BASE COUNT 176 a 159 c 140 g 178 t
ORIGIN
Alignment Scores:
Pred. No.: 104 Length: 984
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x A48781 (1-984)
Qy 176 LeuArgArgIleLeuLeuProLeu 183
|||||
Db 837 CTCCGACGATCCTTCTCTCTA 860

RESULT 81
A58739 A58739 984 bp DNA linear PAT 06-MAR-1998
LOCUS Sequence 2 from Patent WO9700319.
ACCESSION A58739
VERSION A58739.1 GI:3714281
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 984)
AUTHORS Browne,M.J., Chapman,C.G., Clinkenbeard, Helen,E. and Robinson,J.H.
TITLE CHIMERIC LEPTIN FUSED TO IMMUNOGLOBULIN DOMAIN AND USE
JOURNAL Patent: WO 9700319-A 2 03-JAN-1997;
COMMENT Other publication AU 6011096 970115.
FEATURES
Location/Qualifiers
1..984
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 230 a 322 c 269 g 163 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 104 Length: 984
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x A58739 (1-984)

Qy 176 LeuArgArgIleLeuLeuProLeu 183
Db 837 CTCGACGCGATCCTTCTTCCTCTA 860

RESULT 82
AR018707
LOCUS AR018707 984 bp DNA linear PAT 05-DEC-1998
DEFINITION Sequence 8 from patent US 5783181.
ACCESSION AR018707
VERSION AR018707.1 GI:3973821
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 984)
AUTHORS Browne,M.Joseph., Young,P.Ronald., Shatzman,A.Richard.,
Murphy,K.Elizabeth., Chapman,C.Gerald. and
Clinkenbeard,H.Elizabeth.
TITLE Therapeutic uses of fusion proteins between mutant IL 4/IL13
JOURNAL antagonists and immunoglobulins
FEATURES Patent: US 5783181-A 8 21-JUL-1998;
Location/Qualifiers
source 1..984
/organism="unknown"
BASE COUNT 230 a 322 c 269 g 163 t
ORIGIN

Alignment Scores:
Pred. No.: 104 Length: 984
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x AR018707 (1-984)

Qy 176 LeuArgArgIleLeuLeuProLeu 183
Db 837 CTCGACGCGATCCTTCTTCCTCTA 860

RESULT 83
A48778
LOCUS A48778 1006 bp DNA linear PAT 07-MAR-1997
DEFINITION Sequence 5 from patent WO9604388.
ACCESSION A48778
VERSION A48778.1 GI:2302447
KEYWORDS unidentified.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1006)
AUTHORS Browne,M.J., Murphy,K.E., Chapman,C.G., Clinkenbeard,H.E.,
Young,P.R. and Shatzman,A.R.
TITLE NOVEL COMPOUNDS
JOURNAL Patent: WO 9604388-A 5 15-FEB-1996;
SMITHKLINE BEECHAM PLC (GB)
COMMENT Other publication AU 3382595 960304.
FEATURES Location/Qualifiers
source 1..1006
/organism="unidentified"
/db_xref="taxon:32644"

```

```

BASE COUNT 235 a 327 c 271 g 173 t
ORIGIN

Alignment Scores:
Pred. No.: 106 Length: 1006
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x A48778 (1-1006)

Qy 176 LeuArgArgIleLeuLeuProLeu 183
Db 837 CTCGACGCGATCCTTCTTCCTCTA 860

RESULT 84
AR018705
LOCUS AR018705 1006 bp DNA linear PAT 05-DEC-1998
DEFINITION Sequence 5 from patent US 5783181.
ACCESSION AR018705
VERSION AR018705.1 GI:3973819
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1006)
AUTHORS Browne,M.Joseph., Young,P.Ronald., Shatzman,A.Richard.,
Murphy,K.Elizabeth., Chapman,C.Gerald. and
Clinkenbeard,H.Elizabeth.
TITLE Therapeutic uses of fusion proteins between mutant IL 4/IL13
JOURNAL antagonists and immunoglobulins
FEATURES Patent: US 5783181-A 5 21-JUL-1998;
Location/Qualifiers
source 1..1006
/organism="unknown"
BASE COUNT 235 a 327 c 271 g 173 t
ORIGIN

Alignment Scores:
Pred. No.: 106 Length: 1006
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x AR018705 (1-1006)

Qy 176 LeuArgArgIleLeuLeuProLeu 183
Db 837 CTCGACGCGATCCTTCTTCCTCTA 860

RESULT 85
AY094934
LOCUS AY094934 1101 bp mRNA linear INV 15-APR-2002
DEFINITION Drosophila melanogaster RH52423 full insert cDNA.
ACCESSION AY094934
VERSION AY094934.1 GI:20151854
KEYWORDS FLI_CDNA.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,
George,R., Gonzalez,M., Guarin,H., Krommiller,B., Li,P., Liao,G.,
Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Parakas,V., Park,S.,
Patel,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.
and Ceiniker,S.

```

```

source
1. .1149
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT      293 a      344 c      302 g      210 t
ORIGIN

Alignment Scores:
Pred. No.:      119          Length:      1149
Score:          8.00          Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    4.04%          Indels:      0
DB:             6             Gaps:        0

US-09-966-880A-8 (1-198) x A48779 (1-1149)

QY      176 LeuArgArgIleLeuLeuProLeu 183
Db      1002 CTCGACGGATCCTTCTTCTCTA 1025

RESULT 87
LOCUS      A48782          1149 bp      DNA      linear      PAT 07-MAR-1997
DEFINITION Sequence 9 from Patent WO9604388.
ACCESSION  A48782
VERSION    A48782.1 GI:2302450
KEYWORDS   .
SOURCE     unidentified.
ORGANISM   unclassified.
REFERENCE  1 (bases 1 to 1149)
AUTHORS    Browne,M.J., Murphy,K.E., Chapman,C.G., Clinkenbeard,H.E.,
            Young,P.R. and Shatzman,A.R.
TITLE      NOVEL COMPOUNDS
JOURNAL    Patent: WO 9604388-A 9 15-FEB-1996;
            SMITHKLINE BEECHAM PLC (GB)
COMMENT    Other publication AU 3382595 960304.
FEATURES   Location/Qualifiers
            1..1149
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT      293 a      344 c      304 g      208 t
ORIGIN

Alignment Scores:
Pred. No.:      119          Length:      1149
Score:          8.00          Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    4.04%          Indels:      0
DB:             6             Gaps:        0

US-09-966-880A-8 (1-198) x A48782 (1-1149)

QY      176 LeuArgArgIleLeuLeuProLeu 183
Db      1002 CTCGACGGATCCTTCTTCTCTA 1025

RESULT 88
LOCUS      AR018706          1149 bp      DNA      linear      PAT 05-DEC-1998
DEFINITION Sequence 6 from patent US 5783181.
ACCESSION  AR018706
VERSION    AR018706.1 GI:3973820
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 1149)
AUTHORS    Browne,M.Joseph., Young,P.Ronald., Shatzman,A.Richard.,
            Murphy,K.Elizabeth., Chapman,C.Gerald. and
            Clinkenbeard,H.Elizabeth.
            Therapeutic uses of fusion proteins between mutant T1/T113
            and mutant T1/T113
TITLE

```

antagonists and immunoglobulins
JOURNAL Patent: US 5783181-A 6 21-JUL-1998;
FEATURES Location/Qualifiers
source
1. .1149
BASE COUNT 293 a 344 c 302 g 210 t
ORIGIN

Alignment Scores:
Pred. No.: 119 Length: 1149
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x AR018706 (1-1149)
Qy 176 LeuArgArgIleLeuLeuProLeu 183
Db 1002 CTCGACGGATCCTTCTCTCTA 1025

RESULT 89
AR018708
LOCUS AR018708 1149 bp DNA linear PAT 05-DEC-1998
DEFINITION Sequence 9 from patent US 5783181.
ACCESSION AR018708
VERSION AR018708.1 GI:3973822
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1149)
AUTHORS Browne,M.Joseph., Young,P.Ronald., Shatzman,A.Richard.,
Murphy,K.Elizabeth., Chapman,C.Gerald. and
Clinkenbeard,H.Elizabeth.
TITLE Therapeutic uses of fusion proteins between mutant IL 4/IL13
antagonists and immunoglobulins
JOURNAL Patent: US 5783181-A 9 21-JUL-1998;
FEATURES Location/Qualifiers
source
1. .1149
BASE COUNT 293 a 344 c 304 g 208 t
ORIGIN

Alignment Scores:
Pred. No.: 119 Length: 1149
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x AR018708 (1-1149)
Qy 176 LeuArgArgIleLeuLeuProLeu 183
Db 1002 CTCGACGGATCCTTCTCTCTA 1025

RESULT 90
A58738
LOCUS A58738 1188 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 1 from Patent WO9700319.
ACCESSION A58738
VERSION A58738.1 GI:3714280
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1188)
AUTHORS Browne,M.J., Chapman,C.G., Clinkenbeard, Helen,E. and Robinson,J.H.
TITLE CHIMERIC LEPTIN FUSED TO IMMUNOGLOBULIN DOMAIN AND USE
JOURNAL Patent: WO 9700319-A 1 03-JAN-1997;

SMITHKLINE BEECHAM PLC (GB)
COMMENT Other publication AU 6011096 970115.
FEATURES Location/Qualifiers
source
1. .1188
BASE COUNT 282 a 369 c 310 g 227 t
ORIGIN

Alignment Scores:
Pred. No.: 123 Length: 1188
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x A58738 (1-1188)
Qy 176 LeuArgArgIleLeuLeuProLeu 183
Db 1044 CTCGACGGATCCTTCTCTCTA 1067

RESULT 91
A58740
LOCUS A58740 1188 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 3 from Patent WO9700319.
ACCESSION A58740
VERSION A58740.1 GI:3714282
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1188)
AUTHORS Browne,M.J., Chapman,C.G., Clinkenbeard, Helen,E. and Robinson,J.H.
TITLE CHIMERIC LEPTIN FUSED TO IMMUNOGLOBULIN DOMAIN AND USE
JOURNAL Patent: WO 9700319-A 3 03-JAN-1997;
SMITHKLINE BEECHAM PLC (GB)
COMMENT Other publication AU 6011096 970115.
FEATURES Location/Qualifiers
source
1. .1188
BASE COUNT 282 a 369 c 311 g 226 t
ORIGIN

Alignment Scores:
Pred. No.: 123 Length: 1188
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 6 Gaps: 0

US-09-966-880A-8 (1-198) x A58740 (1-1188)
Qy 176 LeuArgArgIleLeuLeuProLeu 183
Db 1044 CTCGACGGATCCTTCTCTCTA 1067

RESULT 92
AF238055
LOCUS AF238055 1230 bp DNA linear PLN 12-JUN-2001
DEFINITION Hydrastis canadensis NADH dehydrogenase subunit F (nuhf) gene,
partial cds; chloroplast gene for chloroplast product.
ACCESSION AF238055
VERSION AF238055.2 GI:10710339
KEYWORDS
SOURCE Hydrastis canadensis.
ORGANISM Chloroplast Hydrastis canadensis
REFERENCE 1 (bases 1 to 1188)
AUTHORS Browne,M.J., Chapman,C.G., Clinkenbeard, Helen,E. and Robinson,J.H.
TITLE CHIMERIC LEPTIN FUSED TO IMMUNOGLOBULIN DOMAIN AND USE
JOURNAL Patent: WO 9700319-A 1 03-JAN-1997;


```

/db_xref="GI:3243077"
/translation="MIWTSARSSFEASLSPSRRTLRLARLAPVEGPTVAELRAALKR
DEFIAPQPVVDAGTGLFGVATTIQNAHRHNGLVPEHCYMAALYRONLHIALAQQLL
AQVFDLAAAMRRNRHVGLEFIQPEPICLSKRSDDVLRNLDHRGLPPEFALTLLAVSPS
LLQSPYQQSLHSURRHAGCNVLTSTFDESPHLDALETCGIALPELMAIYLR
QGGRAACRAIAQAQAQRIPVMAAGVYDEMAQAADLLPCRYLLGDHVPMTGQR
FLYWYRREPPNP"
BASE COUNT      229 a   514 c   461 g   261 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      147      Length:      1465
Score:          8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      4.04%      Indels:      0
DB:              1      Gaps:      0

```

US-09-966-880A-8 (1-198) x AF071567 (1-1465)

```

Qy      126  LeuAqrgrLeuHisArgAlaGly 133
      |||||
Db      1066 CTGCGGCGGTGCACCGGCGGA 1089

```

RESULT 95

EANDHF1

LOCUS

DEFINITION Exacum affine NADH dehydrogenase subunit (ndhf) gene, partial cds;

ACCESSION AF147710

VERSION AF147710.1

KEYWORDS 1 of 2

SEGMENT Exacum affine.

SOURCE Chloroplast Exacum affine

REFERENCE 1 (bases 1 to 1565)

AUTHORS Olmstead,R.G., Kim,K.J., Jansen,R.K. and Wagstaff,S.J.

TITLE The phylogeny of the Asteridae sensu lato based on chloroplast ndhf

JOURNAL Mol. Phylogenet. Evol. 16 (1), 96-112 (2000)

MEDLINE 20340987

PUBMED 10877943

REFERENCE 2 (bases 1 to 1565)

AUTHORS Olmstead,R.G., Kim,K.-J., Jansen,R.K. and Wagstaff,S.J.

TITLE Direct Submission

JOURNAL Submitted (03-MAY-1999) Botany, University of Washington, Campus

Box 355325, Seattle, WA 98195, USA

FEATURES Location/Qualifiers

1..1565

/organism="Exacum affine"

/organelle="plastid:chloroplast"

/db_xref="taxon:13525"

<1..>1565

/gene="ndhf"

/codon_start=2

/product="NADH dehydrogenase subunit"

/protein_id="AA03769.1"

/db_xref="GI:6110322"

/translation="WIIFPLPVPFPGAGLLLPATKSFRRMAFQSVLLISIVM

XFSYLSIQOINSSFFQYVWSMTINNDFSLDFGLDPLTISMSILITVIGTWLIY

SDYNAHQGLRFAFVMSFSTMLGLVTSNNLOIYIFWELVGLCSYLLIGFWPLR

Alignment Scores:

```

Pred. No.:      156      Length:      1565
Score:          8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      4.04%      Indels:      0
DB:              8      Gaps:      0

```

US-09-966-880A-8 (1-198) x EANDHF1 (1-1565)

Qy 42 PheSerLeuAspPheGlyTyrLeu 49

|||||

Db 218 TTTTCTTTAGACTTCGGATACTTG 241

RESULT 96

AF353205

LOCUS

DEFINITION Azospirillum brasilense strain Sp7 poly-beta-hydroxybutyrate

ACCESSION AF353205

VERSION AF353205.1

KEYWORDS GI:21326930

SOURCE Azospirillum brasilense.

ORGANISM Azospirillum brasilense

Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;

Azospirillum.

REFERENCE 1 (bases 1 to 1857)

AUTHORS Kadouri,D., Burdman,S., Jurkevitch,E. and Okon,Y.

TITLE Identification and Isolation of Genes Involved in

Poly(beta-Hydroxybutyrate) Biosynthesis in Azospirillum brasilense

and Characterization of a phbC Mutant

JOURNAL Appl. Environ. Microbiol. 68 (6), 2943-2949 (2002)

MEDLINE 22034968

PUBMED 12039753

REFERENCE 2 (bases 1 to 1857)

AUTHORS Kadouri,D.E., Jurkevitch,E. and Okon,Y.

TITLE Direct Submission

JOURNAL Submitted (25-FEB-2001) Department of Plant Pathology and

Microbiology, The Hebrew University of Jerusalem, Faculty of

Agricultural, Food and Environmental Quality Sciences, Rehovot

76100, Israel

FEATURES Location/Qualifiers

1..1857

/organism="Azospirillum brasilense"

/strain="Sp7"

/db_xref="taxon:192"

1..1857

/gene="phbC"

1..1857

/gene="phbC"

/codon_start=1

/transl_table=11

/product="poly-beta-hydroxybutyrate synthase"

/protein_id="AA048100.1"

/db_xref="GI:21326931"

```

/translation="MVGDLGVDDAQFTORLLDLVLKLPFOVGDAAGDALHFAHLLLLD
LVRVGEERFLSRQADGVGAKNPDPMGVGHAFLEMTTRMADPAKIMKAOMTLWDYF
LTLQWRTTQFFGDAQOPVAPAKDDRRFKDSADWENTLDFDIKOSYLLSARWMSV
NEVDGLDHTAKRVDYFTQFVDAMAFSNFVMTNPEVLTRTTIEGGENLVKGLHLIK
DLERGGELRIKSMTDYDAFOVKGNIAVTGKVVFTDLMLQI1QVTPPTPEVKNRPLMI
LVDALEKVTGKDVAICGCLGCTLLASTLSYMAAKKDDR1KSATFFTMLDFTFAGFL
SVFIDEQLTWIESQAAQOGLGSKMATTFFNMLRANDLIWSFVNNYLLCKDFFPFD
LLYNSDSTRMPAAMHSFYLRNMYOKNLLAOPGAVTLGGVPI1DI.RKVKTPSPFLSARE
DHAPKSTYMGAAHLSFGPYKFLVLAASGHIAGVNPNPAACKYCYWTNAKLPKASDML
ASSEQTPGSMWPEWNNWSTFSECKVPARNPEKGGPLVEDAPGSYAKVRIV"
BASE COUNT      357 a   613 c   575 g   312 t
ORIGIN

```

Alignment Scores:

```

Pred. No.:      181      Length:      1857
Score:          8.00      Matches:      8

```

```

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 1 Gaps: 0

US-09-966-880A-8 (1-198) x AF353205 (1-1857)

QY 123 ProGluGlyLeuArgLeuHis 130 linear PRI 15-JUL-2002
ACCESSION AK094292 1886 bp mRNA BRACE2006249.
VERSION AK094292.1 GI:21753322
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens cerebellum cDNA to mRNA, clone_lib:BRACE2
clone:BRACE2006249.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M.,
Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yanamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kinura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1886)
REFERENCE Isogai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team): 2-6-7-
Kazusa-Kanatairi, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
Location/Qualifiers
1..1886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="BRACE2006249"
/tissue_type="cerebellum"
/clone_lib="BRACE2"
/note="cloning vector: pME18SFL3"
BASE COUNT 540 a 360 c 335 g 651 t
ORIGIN
Alignment Scores:
Pred. No.: 183 Length: 1886
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AK094292 (1-1886)

QY 97 PheLeuArgGlyAsnProAsnLeu 104
DB 1016 TTCTTGAGGGGCAATCCAACTTA 1039

RESULT 98
HAU12660 1938 bp DNA linear PLN 19-APR-2002
LOCUS Hemigraphis alternata NADH dehydrogenase subunit (ndhF) gene,
DEFINITION partial cds; chloroplast gene for chloroplast product.
ACCESSION U12660
VERSION U12660.1 GI:607930
SOURCE Hemigraphis alternata.
ORGANISM Chloroplast Hemigraphis alternata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Lamiales; Acanthaceae; Hemigraphis.
1 (bases 1 to 1938)
REFERENCE Scotland,R.W., Sweere,J.A., Reeves,P.A. and Olmstead,R.G.
TITLE Higher-level systematics of Acanthaceae determined by chloroplast
DNA sequences
JOURNAL Am. J. Bot. 82 (2), 266-275 (1995)
AUTHORS 2 (bases 1 to 1938)
REFERENCE Olmstead,R.G.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-1994) Richard G. Olmstead, E.P.O. Biology,
University of Colorado, Boulder, CO 80309, USA
FEATURES
source
Location/Qualifiers
1..1938
/organism="Hemigraphis alternata"
/organelle="plastid:chloroplast"
/db_xref="taxon:34269"
1..1938
/gene="ndhF"
/gene="ndhF"
/note="151 nucleotides from the PCR primer near the 5' end
were unsequenced"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase subunit"
/protein_id="AAA61724.1"
/db_xref="GI:607931"
VGI:AVLYSDNYMAHDGYLRFPAYMSFFSTMLGLVTSNLIQIYIPWELVGMCSYL
LIGWFTPLAADCQAKAFV7NRVGDGFLGILFCLFWITGSFRDLUFEMFNNSYIN
NQVNFPTLCVLLFAGAVAKSAQPLHWLPDAMEGPTISALIHAATVWAAGVFL
VARLLPFWIPYIMDFITSLIGIITFLGATLALAKDKIKRGLAYSTMQLGYMMAL
ITKISFLLGTFSLCGVPLACFWKDEILNDTWLYSPIAIIATAGLTXYMFRIY
LLTEGHLNVHFQNSANQIFLYISLWACSKKININERLLRIKNTSFSLKPTP
ORDEKINRPFINILSFHNKTFSTPESDNTMLFSLLLILVLTFSFVGSIGIPNPF
NQEGTDILSKWVTPTSIHLHQAEKSTHWYEFKDATFSVSLAYFGIFLAALLKPIY
SSFNFDLVNSFVKTNGLNRDWDKVINALYNWYNRAYIDVFYRTSVIAMIRGLAKLT
HFVDRRIIDGITNLVGLLSXPI"
BASE COUNT 535 a 303 c 342 g 756 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 187 Length: 1938
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 8 Gaps: 0

US-09-966-880A-8 (1-198) x HAU12660 (1-1938)

QY 42 PheSerLeuAspPheGlyTyrLeu 49
DB 67 TTTCTTTAGACTTCGGCTACTTG 90

RESULT 99
AF257485 1990 bp DNA linear PLN 11-JUN-2000
LOCUS Goyazia rupicola NADH dehydrogenase (ndhF) gene, partial cds;
DEFINITION chloroplast gene for chloroplast product.

```

```

ACCESSION AF257485 GI:8453157
VERSION AF257485.1
KEYWORDS
SOURCE
ORGANISM
Chloroplast Goyazia rupicola
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Asteridae; euasterids I; Lamiales; Gesneriaceae; Goyazia.
REFERENCE 1 (bases 1 to 1990)
AUTHORS Smith,J.F.
TITLE The phylogenetic relationships of Lemnocarpaceae and Goyazia
(Gesneriaceae) based on ndhF sequences
JOURNAL Ann. Mo. Bot. Gard. (2000) In press
REFERENCE 2 (bases 1 to 1990)
AUTHORS Smith,J.F.
TITLE Direct Submission
JOURNAL Submitted (18-APR-2000) Biology, Boise State University, 1910
University Drive, Boise, ID 83725, USA
FEATURES
source
1..1990
Location/Qualifiers
/organism="Goyazia rupicola"
/organelle="plastid:chloroplast"
/db_xref="taxon:125973"
<1..>1990
/gene="ndhF"
<1..>1990
/gene="ndhF"
/codon_start=1
/product="NADH dehydrogenase"
/protein_id="AAF75262.1"
/db_xref="GI:8453158"
/translation="GMIIPILIPVPMLIGTGLLLPPTATKRLRMALPSILLSTV
MFSINLSIOLINSISIIYQVSWTNDXDFSLDFGLIDLLTSMSLITTVGIMVLI
YSDNYMAHDOGYLRLEAYMSEFSPMGFLVTSNLIQILFELWELVGMCSVILIGFWET
RPLAANCQKAFVYNNVGDRLGLLGISGFWIGSGFEDRLFEILLNINNEVDFFP
VTLCAALLFAGAVAKSVQFPLHWLPDAMEGPTDISALIHAATWVAAGIFLVARLPL
FIATPYIMYSVVGIIITVLGATLALAKQDKRGLAYSAMSOQLGYMMLALMGSYRC
ALPHLITHAYSKALLFLGSGSIHSMETIVGYSQKQNMGLMGGLTKHVPITKISPL
FGTSLGCIPLPACFWKSKDELNDWSLSPFIATASATAGTAVTAFVRIYLLTFEGH
LNIHFQNSKQNSPIYSISITWGXKXKXKRLNDNNEXXTFFFAKTKRSDKGKIK
MARPFIIVHFDKPKSPYIESDNTMLFLLVLVLTFLVGLGIFPNXXEXKLDLIL
TKWLAPSNILHOKNSDMOWIEFFKGAVASVSIAYXGIFVASFLYKPIYSSSKNFEL
INSFVKGQPSILMDKINGIYDMSYNRAVIDTFYTRSF"
BASE COUNT 542 a 313 c 346 g 770 t 19 others
ORIGIN
Alignment Scores:
Pred. No.: 192 Length: 1990
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.04% Indels: 0
DB: 8 Gaps: 0
US-09-966-880A-8 (1-198) x AF257485 (1-1990)
Qy 42 PheSerLeuAspPheGlyTyrLeu 49
|||||
Db 220 TTTTCTTTAGATTTCGGCTACTTC 243
RESULT 100
LOCUS ESU12659
DEFINITION Eremomastax speciosa NADH dehydrogenase subunit (ndhF) gene,
partial cds; chloroplast gene for chloroplast product.
ACCESSION UI2659
VERSION UI2659.1 GI:607928
KEYWORDS
SOURCE
Eremomastax speciosa.
Chloroplast Eremomastax speciosa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Lamiales; Acanthaceae; Eremomastax.

```

```

REFERENCE 1 (bases 1 to 2025)
AUTHORS Scotland,R.W., Sweere,J.A., Reeves,P.A. and Olmstead,R.G.
TITLE Higher-level systematics of Acanthaceae determined by chloroplast
DNA sequences
JOURNAL Am. J. Bot. 82 (2), 266-275 (1995)
REFERENCE 2 (bases 1 to 2025)
AUTHORS Olmstead,R.G.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-1994) Richard G. Olmstead, E.P.O. Biology,
University of Colorado, Boulder, CO 80309, USA
FEATURES
source
1..2025
Location/Qualifiers
/organism="Eremomastax speciosa"
/organelle="plastid:chloroplast"
/db_xref="taxon:37794"
1..2025
/gene="ndhF"
<1..>2025
/gene="ndhF"
/note="70 nucleotides from the PCR primer near the 5' end
were unsequenced"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase subunit"
/protein_id="AA61723.1"
/db_xref="GI:607929"
/translation="ATKNRLRLWLTPTXXSLTVMIFSTKLSIEQINGSSIIYQYVXW
TFDNDFSLDFGLIDPLTSLIMLITTVGIIITVLIYSDNYMAHDOGYLRFPAYMSFCT
AMGLVTSNLIQIYIFWELVGMCSVILIGFWETRPLAADACQKAFVNNRVGDFGLLL
GIFCFWITGTFEFDRLFEFNNSVYNNVNOVFPVTLCTVLLFAGAVAKSAQFPLHW
LPDAMEGPTDISALIHAATWVAAGVFLVARLLPFVPIVIMDIISIIGITILLGAT
LALAQDKIRGLAYSTMOLGYMMLALMGCSRSALPHLITHAYSKALLFLGSCSVIH
SMESIVGYSQKQNMGLMGGLTKHLPITKIAELLGTFLGCGVPPPLACFWKSKDEILNH
TWLSPSFALIASITAGLTAFYMFRIYLLTTFEGHNVHVLQNYSANONIFLYISIMCK
ACSKRINFNFSLRIYNTSSSLKKTQORDEKRRGNWPF1KILYSRKNQTFESYKE
SDNTMLFSLILVLTFLVGSIGIPFNENQEGTDILSKWVTDSIILHQKAKDSTHW
YEFEKDATFSVLAYTGIFLASSLYKPIYSSKFNFDLINSFVKTGLNRRNDKMINVL
YNNWNRGYIDTFYITTSFFAMIRGLAKLTHFVDRRIIDGITNVVGLMSPL"
BASE COUNT 557 a 310 c 360 g 788 t 10 others
ORIGIN
Alignment Scores:
Pred. No.: 195 Length: 2025
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0
US-09-966-880A-8 (1-198) x ESU12659 (1-2025)
Qy 42 PheSerLeuAspPheGlyTyrLeu 49
|||||
Db 148 TTTTCTTTAGACTTCGGTTACTTC 171
Search completed: July 17, 2003, 22:23:21
Job time : 2445 secs

```